INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Moments and Tails in Monotone–Separable Stochastic Networks

François Baccelli — Serguei Foss

N° 4197

June 2001

THÈME 1

ISSN 0249-6399 ISRN INRIA/RR--4197--FR+ENG

INRIA
ROCQUENCOURT

Moments and Tails in Monotone–Separable Stochastic Networks

François Baccelli * , Serguei Foss †

Thème 1 — Réseaux et systèmes
Projet TREC

Rapport de recherche n° 4197 — June 2001 — 40 pages

Abstract: A network belongs to the monotone separable class if its state variables are homogeneous and monotone functions of the epochs of the arrival process. This framework, which was first introduced to derive the stability region for stochastic networks with stationary and ergodic driving sequences, is revisited. It contains several classical queueing network models, including generalized Jackson networks, max plus networks, multiserver queues, and various classes of stochastic Petri nets. Our purpose is the analysis of the tails of the stationary state variables in the particular case of i.i.d. driving sequences. For this, we establish general comparison relationships between networks of this class and the GI/GI/1/∞ queue. We first use this to show that two classical results of the asymptotic theory for GI/GI/1/∞ queues can be directly extended to this framework. The first one concerns the existence of moments for the stationary state variables. We establish that for all $\alpha \geq 1$, the $\alpha+1$-moment condition for service times is necessary and sufficient for the existence of the α-moment for the stationary maximal dater (typically the time to empty the network when stopping further arrivals) in any network of this class. The second one is a direct extension of Veraverbeke's tail asymptotic for the stationary waiting times in the GI/GI/1/∞ queue. We show that under sub-exponential assumptions for service times, the stationary maximal dater in any such network has tail asymptotics which can be bounded from below and from above by a multiple of the second tails of service times. In general, the upper and the lower bounds do not coincide. Nevertheless, exact asymptotics can be obtained along the same lines for various special cases of networks, providing direct extensions of Veraverbeke's tail asymptotic for the stationary waiting times in the GI/GI/1/∞ queue. We exemplify this on tandem queues (maximal daters and delays in stations) as well as on multiserver queues. This methodology for exact asymptotics can be extended to other classes of monotone separable networks like general reducible max plus networks, or generalized Jackson networks.

* INRIA-ENS, ENS, 45 rue d'Ulm 75005 Paris, France {Francois.Baccelli@ens.fr} The work of this author was supported by the TMR Alapedes project and by INTAS.

† Institute of Mathematics, 630090 Novosibirsk, Russia & Department of Actuarial Mathematics and Statistics, Heriot-Watt University, Edinburgh, UK. {S.Foss@ma.hw.ac.uk} The work of this author was supported by INTAS and by the Lyapunov Center.

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau, Rocquencourt, BP 105, 78153 Le Chesnay Cedex (France)
Téléphone : +33 1 39 63 55 11 — Télécopie : +33 1 39 63 53 30

Key-words: Open queueing network, ergodic property, sub-additive ergodic theorem, sub-exponential random variable, heavy tail, integrated tail, Veraverbeke's theorem, Jackson network, max plus network, tandem queue, multiserver queue.

Moments et Queues des Distributions de Réseaux Stochastiques Monotones et Séparables

Résumé : Un réseau ouvert appartient à la classe monotone-séparable si ses variables d'état sont des fonctions homogènes et monotones des dates d'arrivées dans le réseau. Ce cadre, qui a initialement été introduit pour analyser la région de stabilité de réseaux stochastiques sous des hypothèses stationnaires ergodiques, contient plusieurs modèles classiques comme les réseaux de Jackson généralisés, les réseaux (max,plus) linéaires, les files multiserveurs et diverses classes de réseaux de Petri stochastiques. Le but de cet article est l'analyse des queues des distributions stationnaires des variables d'état de ces réseaux dans le cas i.i.d. Pour ce faire, nous établissons deux relations de comparaison entre les réseaux de cette classe et la file d'attente GI/GI/1/∞. Nous utilisons ces relations pour généraliser deux résultats asymptotiques bien connus sur l'état stationnaire de la file d'attente GI/GI/1/∞ aux réseaux monotones-séparables. Le premier d'entre eux concerne les moments des variables d'état stationnaires. Nous montrons que pour tout $\alpha \geq 1$, l'hypothèse de moment d'ordre $\alpha + 1$ pour les temps de service est nécessaire et suffisante pour l'existence du moment d'ordre α du dateur maximal stationnaire (temps pour vider le réseau quand on arrête les arrivées) de tout réseau de cette classe. Le second résultat est une extension directe du théorème de Veraverbeke sur le comportement asymptotique de la queue de la distribution du temps d'attente stationnaire dans la file GI/GI/1/∞. Nous montrons que sous des hypotèses sous-exponentielles sur les temps de service, le dateur maximal stationnaire de tout réseau de cette classe a une distribution dont la queue peut être asymptotiquement majorée et minorée par des multiples de la queue de la distribution d'excès des services. En général, ces deux bornes ne coïncident pas. Néanmoins, des asymptotiques exactes peuvent être obtenues dans la lignée des résultats précédents pour divers cas particulier de réseaux. Nous donnons en particulier des extensions du théorème de Veraverbeke pour des files d'attente en série (dateur maximal et délais dans les stations) ainsi que pour les files multiserveurs. Cette méthode pour obtenir les asymptotiques exactes se généralise à d'autres classes de réseaux monotones-séparables comme les réseaux (max,plus) linéaires généraux (irréductibles ou non) ou encore les réseaux de Jackson généralisés.

Mots-clés : réseau ouvert, file d'attente, ergodicité, sous-additivité, sous-exponentialité, variable aléatoire, queue épaisse, distribution d'excès, théorème de Veraverbeke, réseau de Jackson, réseau (max,plus) linéaire, files en série, file multiserveur.

1 Introduction

We show in the present paper that properties which have been known for a long time for the tail asymptotics of isolated single server queues can be extended to the class of stochastic networks which are *monotone and separable.*

Section 2 summarizes the definition and main results on monotone–separable networks, and in particular the ergodic theorems that allow one to determine their stability region.

Section 3 focuses on the proof of the moment theorem. The assumptions that are needed here are limited to independence. In particular, no Harris chain representation that would impose specific assumptions on service times is required.

The sub-exponential tail asymptotic theorems are given in §4 and 5. For surveys on the state of the art for this kind of asymptotics, see [13].

Section 4 gives generic upper and lower bounds which hold for all monotone separable networks, and which only differ in the multiplicative constants.

Section 5 focuses on more precise asymptotics for specific cases. The key idea, which is based on a corollary of Veraverbeke's theorem and on the bounds established in §4, is summarized in §5.2. In the present paper, we limit ourselves to two simple examples: tandem queues (which are both a special instance of generalized Jackson network and a special instance of reducible max plus network) and multiserver queues. In §5.3.1 and 5.3.2, we obtain the exact asymptotics for tandem queues. In §5.3.3 multiserver queues are also investigated within this framework.

To the best of our knowledge, among the various classes of networks listed above, exact asymptotics are only known for irreducible max plus networks [6]. The results on tandem queues can be extended to more general reducible max plus networks. Sharp asymptotics for generalized Jackson networks can also be obtained along the same lines. This will be the object of a companion paper [4].

2 Basic Results on the Monotone Separable Networks

2.1 Framework

Consider a stochastic network described by the following framework:

- The network has a single input point process N, with points $\{T_n\}$; for all $m \leq n \in \mathbb{N}$, let $N_{[m,n]}$ be the $[m,n]$ restriction of N, namely the point process with points $\{T_l\}_{m \leq l \leq n}$.
- The network has a.s. finite activity for all finite restrictions of N: for all $m \leq n \in \mathbb{N}$, let $X_{[m,n]}(N)$ be the time of the last activity in the network, when this one starts empty and is fed by $N_{[m,n]}$. We assume that for all finite m and n as above, $X_{[m,n]}$ is finite.

We assume that there exists a set of functions $\{f_l\}$, $f_l : \mathbb{R}^l \times K^l \to \mathbb{R}$, such that:

$$X_{[m,n]}(N) = f_{n-m+1}\{(T_l, \xi_l),\ m \leq l \leq n\}, \tag{1}$$

for all n, m and N, where the sequence $\{\xi_n\}$ is that describing service times and routing decisions.

We say that a network described as above is monotone–separable if the functions f_n are such that the following properties hold for all N:

1.**(causality):** for all $m \leq n$,

$$X_{[m,n]}(N) \geq T_n;$$

2.**(external monotonicity):** for all $m \leq n$,

$$X_{[m,n]}(N') \geq X_{[m,n]}(N),$$

whenever $N' \stackrel{\text{def}}{=} \{T'_n\}$ is such that $T'_n \geq T_n$ for all n, a property which we will write $N' \geq N$ for short;

3.**(homogeneity):** for all $c \in \mathbb{R}$ and for all $m \leq n$

$$X_{[m,n]}(c + N) = X_{[m,n]}(N) + c;$$

4.**(separability):** if, for all $m \leq l < n$, $X_{[m,l]}(N) \leq T_{l+1}$, then

$$X_{[m,n]}(N) = X_{[l+1,n]}(N).$$

2.2 Maximal Daters

By definition, the $[m, n]$ maximal dater is

$$Z_{[m,n]}(N) \stackrel{\text{def}}{=} X_{[m,n]}(N) - T_n = X_{[m,n]}(N - T_n).$$

Note that $Z_{[m,n]}(N)$ is a function of $\{\xi_l\}$ and $\{\tau_l\}_{m \leq l \leq n-1}$ only, where $\tau_n = T_{n+1} - T_n$. In particular, $Z_n(N) \stackrel{\text{def}}{=} Z_{[n,n]}(N)$ is not a function of $\{\tau_n\}$.

Lemma 1 (internal monotonicity of X and Z) *Under the above conditions, the variables $X_{[m,n]}$ and $Z_{[m,n]}$ satisfy the internal monotonicity property: for all N*

$$\begin{aligned} X_{[m-1,n]}(N) &\geq X_{[m,n]}(N), \\ Z_{[m-1,n]}(N) &\geq Z_{[m,n]}(N), \ (m \leq n). \end{aligned}$$

Lemma 2 (sub-additive property of Z) *Under the above conditions, $\{Z_{[m,n]}\}$ satisfies the following sub-additive property: for all $m \leq l < n$, for all N*

$$Z_{[m,n]}(N) \leq Z_{[m,l]}(N) + Z_{[l+1,n]}(N).$$

2.3 Stochastic Assumptions

Assume the variables $\{\tau_n, \xi_n\}$ are random variables defined on a common probability space $(\Omega, \mathcal{F}, \mathbf{P}, \theta)$, where θ is an ergodic, measure-preserving shift transformation, such that $(\tau_n, \xi_n) \circ \theta = (\tau_{n+1}, \xi_{n+1})$. The following integrability assumptions are also assumed to hold:

$$\mathbf{E}\tau_n \stackrel{\text{def}}{=} \lambda^{-1} \stackrel{\text{def}}{=} a < \infty, \quad \mathbf{E}Z_n < \infty.$$

Most of the new results of the present paper will be obtained under the following independence assumption:

(IA): *the sequences $\{\xi_n\}$ and $\{\tau_n\}$ are mutually independent and each of them consists of i.i.d. random variables.*

2.4 Main Stability Results

We summarize the main results of [3]. These results do not require (IA).

2.4.1 First-Order Ergodic Property

Kingman's sub-additive ergodic theorem gives:

Lemma 3 *There exists a finite non-negative constant γ such that the a.s. limits*

$$\lim \frac{Z_{[-n,-1]}}{n} = \lim \frac{\mathbf{E}Z_{[-n,-1]}}{n} = \lim \frac{Z_{[1,n]}}{n} = \lim \frac{\mathbf{E}Z_{[1,n]}}{n} = \gamma$$

hold $\mathbf{P}$*-a.s.*

Corollary 1 *Under the foregoing assumptions*

$$\lim_n \frac{X_{[1,n]}}{n} = \gamma + \lambda^{-1}.$$

2.4.2 0-1 Law

Let A be the event $A = \{\lim Z_{[-n,0]} = \infty\}$.

Theorem 1 *Under the foregoing ergodic assumption,* $\mathbf{P}(A) \in \{0, 1\}$.

2.4.3 Scaling Factor

For all $0 \leq c < \infty$, the sequences

$$X_{[m,n]}(cN) \stackrel{\text{def}}{=} f_{n+1-m}\{(c \cdot T_l, \xi_l); m \leq l \leq n\}$$

and

$$Z_{[m,n]}(cN) = X_{[m,n]}(cN) - c \cdot T_n$$

satisfy all the monotonicity and sub-additive properties mentioned above. In addition, for all n

(a) $Z_{[-n,-1]}(cN)$ is decreasing in c;

(b) $X_{[1,n]}(cN)$ is increasing in c.

Thus

Lemma 4 *For all $c \geq 0$, there exists a non-negative constant $\gamma(c)$ such that*

$$\lim \frac{Z_{[-n,-1]}(cN)}{n} = \gamma(c) \quad \text{a.s.};$$

$\gamma(c)$ is decreasing in c while $\gamma(c) + c\lambda^{-1}$ is increasing in c. In particular, $\gamma + \lambda^{-1} \geq \gamma(0)$, where $\gamma \stackrel{\text{def}}{=} \gamma(1)$.

2.4.4 Second-Order Ergodic Property

The main result on the stability region is:

Theorem 2 *If* $\lim Z_{[-n,0]}(N) = \infty$ *a.s., then* $\lambda\gamma(0) \geq 1$. *If* $\lambda\gamma(0) > 1$, *then* $\lim Z_{[-n,0]}(N) = \infty$ *a.s.*

2.5 Further Assumptions

Let Q the point process with all its points equal to 0: $T_n(Q) = 0$ for all n. For certain results, we shall make the following additional assumption:

(AA)

- *For all* i,
$$Z_i = Z_{[i,i]} = Y_i^{(1)} + \cdots + Y_i^{(r)}, \tag{2}$$
where the r.v.'s $Y_i^{(j)}$ *are non negative, independent of inter-arrival times, and such that the sequence of random vectors* $(Y_i^{(1)}, \ldots, Y_i^{(r)})$ *is i.i.d; general dependences between the coordinates of the vector* $(Y_i^{(1)}, \ldots, Y_i^{(r)})$ *are allowed.*

- *In addition,*
$$Z_{[n,0]}(Q) \geq \max_{j=1,\ldots,r} \sum_{i=n}^{0} Y_i^{(j)} \quad a.s. \tag{3}$$

2.6 Upper and Lower Bound G/G/1/∞ Queues

The results of this section are new. We assume stability, namely $\gamma(0) < a$. We pick an integer $L \geq 1$ such that
$$\mathbf{E} Z_{[-L,-1]}(Q) < La, \tag{4}$$
which is possible in view of Theorem 2. Without loss of generality, one can assume $T_0 = 0$.

To the input process N, we associate the following lower and upper bound processes: $N^- = \{T_n^-\}$, where, for all k and n in $\mathbb{Z}$ such that $n = (k-1)L + 1, \ldots, kL$, $T_n^- = T_{(k-1)L}$, and similarly, $N^+ = \{T_n^+\}$, where $T_n^+ = T_{kL}$ if $n = (k-1)L + 1, \ldots, kL$. Then for all n
$$X_{[-n,0]}(N^-) \leq X_{[-n,0]}(N) \leq X_{[-n,0]}(N^+) \equiv Z_{[-n,0]}(N^+). \tag{5}$$

Note that if (IA) holds, the r.v.'s $Z_{[-n,0]}(N^-) = X_{[-n,0]}(N^-) - T_{-L}$ and $Z_{[-n,0]}(N^+)$ have the same distribution, and that the r.v.'s $Z_{[-n,0]}(N^-)$ and T_{-L} are independent.

2.6.1 Upper Bound Queue

The next lemma, which establishes a first connection between monotone-separable networks and the G/G/1/∞ queue, directly follows from the monotonicity and the separability assumptions.

Lemma 5 *Assume* $T_0 = 0$. *For any* $m < n \leq 0$,
$$Z_{[m,0]}(N) \leq Z_{[n,0]}(N) + \max(0, Z_{[m,n-1]}(N) - \tau_{n-1}).$$

Put $Z_n = Z_{[n,n]}(N)$. Then the sequence $\{Z_n\}$ does not depend on N and forms a stationary and ergodic sequence.

Corollary 2 *Assume $T_0 = 0$. For any $m < 0$,*

$$Z_{[m,0]}(N) \leq \max_{m \leq k \leq 0} \left(\sum_{i=k}^{0} Z_i - \sum_{i=k+1}^{0} \tau_i \right),$$

with the convention $\sum_1^0 = 0$.

The main weakness of this upper bound comes from the fact that the corresponding queue may be unstable whereas the initial network is stable. This is taken care of by the upper bound described below.

Corollary 3 *The stationary maximal dater $Z_{[-\infty,0]} \equiv Z_{[-\infty,0]}(N)$ is bounded from above by the stationary response time $\widehat{R}$ in the $G/G/1/\infty$ queue with service times*

$$\widehat{s}_n = Z_{[L(n-1)+1,Ln]}(Q) \tag{6}$$

and inter-arrival times $\widehat{\tau}_n = T_{Ln} - T_{L(n-1)}$, where L is the integer defined in (4). Since $\widehat{b} = \mathbf{E}\widehat{s}_n < \mathbf{E}\widehat{\tau}_n = La$, this queue is stable.

Proof We have

$$\begin{aligned}
Z_{(-\infty,0]} &= \lim_{n\to\infty} Z_{[-n,0]} \\
&= \lim_{k\to\infty} Z_{[-kL+1,0]} \\
&= \sup_{k\geq 1} Z_{[-kL+1,0]} \\
&\leq \sup_{k\geq 1} Z_{[-kL+1,0]}(N^+) \\
&\leq \sup_{k\geq 1} \max_{-k\leq i\leq 0} \left(\widehat{s}_0 + \sum_{k=i}^{(-1)} (\widehat{s}_j - \widehat{\tau}_{j+1}) \right) \\
&= \sup_{k\geq 1} \left(\widehat{s}_0 + \sum_{i=-k}^{(-1)} (\widehat{s}_i - \widehat{\tau}_{i+1}) \right) = \widehat{R}.
\end{aligned}$$

In these relations, (5) was used to derive the first inequality, Corollary 2 was used in the last inequality; we also used the fact that

$$Z_{[L(n-1)+1,Ln]}(N^+) = Z_{[L(n-1)+1,Ln]}(Q)$$

and the convention $\sum_0^{-1} = 0$. □

The queue of Corollary 3 will be referred to as the L–upper–bound G/G/1/∞ queue associated with the network.

Note that when (IA) holds, this queue is a GI/GI/1/∞ queue. In this case, $\widehat{R} = \widehat{W} + \widehat{s}_0$, where $\widehat{W}$ is a stationary waiting time and $\widehat{W}$ and $\widehat{s}_0$ are independent.

Notice that under (AA),

$$\max_{j=1,\ldots,r} \sum_{i=L(n-1)+1}^{Ln} Y_i^{(j)} \leq \widehat{s}_n \leq \sum_{j=1}^{r} \sum_{i=L(n-1)+1}^{Ln} Y_i^{(j)} \tag{7}$$

a.s. where the second inequality follows from the sub-additive property of Z.

2.6.2 Lower Bound Fork-Join Queue

The following result is immediate:

Lemma 6 *Under Condition (AA),*

$$Z_{(-\infty,0]} \quad \geq \quad \underline{R} = \max_{j=1,\ldots,r} \sup_{n\leq 0} \left(\sum_{n}^{0} Y_i^{(j)} - \sum_{n}^{-1} \tau_i \right). \tag{8}$$

The queue with service times $\{Y_i^{(j)}\}$ and interarrival times $\{\tau_i\}$ will be referred to as the j–lower–bound G/G/1/∞ queue associated with the network. Let $R^{(j)}$ denote the stationary response time in this queue:

$$R^{(j)} = \sup_{n\leq 0} \left(\sum_{n}^{0} Y_i^{(j)} - \sum_{n}^{-1} \tau_i \right).$$

Then the lower bound $\underline{R}$ defined in (8) is the stationary response time in the r-dimensional *fork join queue* with service times $\{Y_i^{(j)}\}$, $j = 1, \ldots, r$ and interarrival times $\{\tau_i\}$.

2.7 Examples

2.7.1 Tandem Queues

Consider a stable ./GI/1/ ∞ tandem queue. Denote $\{\sigma_n^{(i)}\}$ the sequence of service times in station $i = 1, 2$ and $\{\tau_n\}$ the sequence of inter-arrival times at the first station. Put $b^{(i)} = \mathbf{E}\sigma^{(i)}$, $a = \mathbf{E}\tau$, and $\rho^{(i)} = \frac{b^{(i)}}{a} < 1$. We have $\gamma(0) = \max(b^{(1)}, b^{(2)})$.

Tandem queues fall in the class of open Jackson networks, and in the class of open max-plus systems which both belong to the class of monotone separable networks (see below). We have the following representation for the maximal dater (see e.g. [6]),

$$Z_{[-n,0]} \quad = \quad \sup_{-n\leq p\leq 0} \quad \sup_{p\leq q\leq 0} \left(\sum_{m=p}^{q} \sigma_m^{(1)} + \sum_{m=q}^{0} \sigma_m^{(2)} - (T_0 - T_p) \right) \tag{9}$$

$$Z \quad = \quad Z_{[-\infty,0]} = \sup_{p\leq 0} \quad \sup_{p\leq q\leq 0} \left(\sum_{m=p}^{q} \sigma_m^{(1)} + \sum_{m=q}^{0} \sigma_m^{(2)} - (T_0 - T_p) \right). \tag{10}$$

Assumption (IA) is satisfied under the above independence assumptions. Assumption (AA) is also satisfied here with $r = 2$ and $Y_n^{(i)} = \sigma_n^{(i)}$, $i = 1, 2$.

The maximal dater associated with customer n is the sojourn time of this customer in the network, namely the time which elapses between its arrival in station 1 and its departure from station 2.

As for the L–upper–bound queue associated with this network, the expression for $\widehat{s}_n$ is here

$$\widehat{s}_n = \max_{1\le j\le L}\left(\sum_{i=1}^{j}\sigma^{(1)}_{(n-1)L+i} + \sum_{i=j}^{L}\sigma^{(2)}_{(n-1)L+i}\right). \tag{11}$$

2.7.2 Multiserver Queues

Let

$$W_n = (W_n^{(1)},\dots,W_n^{(m)})$$

be the Kiefer-Wolfowitz workload vector in the GI/GI/m/∞ queue with inter-arrival times $\{\tau_n\}$ and service times $\{\sigma_n\}$. Here n is the customer index and $W_n^{(i)}$, $i=1,\dots,m$ are the workloads of the servers at the n-th arrival time, arranged in a non-decreasing order. More precisely, we assume $W_0=(0,\dots,0)$ and

$$W_{i+1} = \mathbf{R}(W_n + \mathbf{e}_1\sigma_i - \mathbf{i}\tau_i)^+ \tag{12}$$

for $i\ge 0$, where $\mathbf{e}_1=(1,0,\dots,0)$ and $\mathbf{i}=1,1,\dots,1)$ are m-dimensional vectors and the operator $\mathbf{R}$ permutes the coordinates of a vector in the non-decreasing order. For a multi-server queue, $\gamma(0)=\frac{\mathbf{E}\sigma_0}{m}$.

Assumption (IA) is satisfied under the assumption that the service times are i.i.d. Assumption (AA) is not satisfied here.

The maximal dater associated with customer n is the time which elapses between its arrival the time when all customers still present at its arrival time have left the system (including customer n):

$$Z_{[0,n]} = \max(W_n^{(1)}+\sigma_n, W_n^{(m)}).$$

2.7.3 Generalized Jackson Networks

Consider a generalized Jackson network with r stations. We denote

- $\{\sigma_n^{(k)}\}$ the i.i.d. sequence of service times in station k;
- $\{\nu_n^{(i)}\}$ the i.i.d. sequence of routing decisions from station i (with values in the set $\{1,\dots,r\}$;
- $\{\nu_n\}$ the i.i.d. sequence of routing decisions for the input process (also with values in the set $\{1,\dots,r\}$;
- $\{\tau_n\}$ the i.i.d. sequence of interarrival time.

Under these assumptions, both (IA) and (AA) are satisfied. We have:

$$Y_1^{(j)} = \sum_{1}^{N(j)}\sigma_n(j) \tag{13}$$

with $N(j)$ the total number of visits of customer 1 (the customer arriving at time T_1) to station j in the $[1,1]$ restriction of the network, namely when this customer is the only one to

enter the network. The random variables $N(j)$, $j = 1, \ldots, r$ are obtained from the sequences of routing decisions (see [2]).

In this case $Z_{[-n,0]}$ is the time which elapses between the arrival of customer 0 and the time when all customers have left the system, given that arrivals are stopped after T_0.

2.7.4 Max Plus Networks

The class of open max plus networks also falls in this framework (see e.g. [3]). A typical example of this class is that of tandem queues. Tandem queues form a reducible open max plus network. For examples of irreducible networks of this class, see [6].

3 Integrability of Stationary Maximal Daters

We assume (IA) and stability, namely $\lambda\gamma(0) < 1$.

3.1 Upper Bound

Let $\widehat{W}$ denote the stationary waiting time in the L–upper–bound GI/GI/1/∞ queue of the network. The following result is well-known

Lemma 7 *For any $\alpha > 1$, $\mathbf{E}\widehat{W}^{\alpha-1}$ is finite if and only if $\mathbf{E}\widehat{s}_0^{\alpha}$ is finite.*

Therefore, $\widehat{R} = \widehat{W} + \widehat{s}$ is such that $\mathbf{E}\widehat{R}^{\alpha-1}$ is finite if and only if $\mathbf{E}\widehat{s}_0^{\alpha}$ is finite.

Corollary 4 *If $\mathbf{E}Z_0^{\alpha} < \infty$, then $\mathbf{E}Z_{(-\infty,0]}^{\alpha-1} < \infty$.*

Proof We have

$$\widehat{s}_0 \leq \sum_{-L+1}^{0} Z_i.$$

Therefore if $\mathbf{E}Z_0^{\alpha} < \infty$, then $\mathbf{E}\widehat{s}_0^{\alpha}$ is finite. Thus, $\mathbf{E}\widehat{W}^{\alpha-1}$ and $\mathbf{E}\widehat{R}^{\alpha-1}$ are finite, too. We conclude the proof by using the bound $Z_{[-\infty,0]} \leq \widehat{R}$ (see the proof of Corollary 3). □

3.2 Lower Bound

Under condition (AA), $\mathbf{E}Z_0^{\alpha}$ is finite if and only if for all j, $\mathbf{E}\left[(Z_0(j))^{\alpha}\right]$ is finite. The following lemma is then an immediate consequence of Lemma 6 and Lemma 7.

Lemma 8 *Under assumptions (IA) and (AA), if $\mathbf{E}\left[Z_{(-\infty,0]}^{\alpha-1}\right]$ is finite, then $\mathbf{E}Z_0^{\alpha}$ is finite too.*

3.3 Examples

All results are given are under the assumption that the system under consideration is stable.

- Tandem Queues: the system response time has a moment of order α iff the service times in both stations admit a moment of order $\alpha - 1$. This property holds true for both the case when service times of the two stations are independent, and the case when the n-th service times are identical.

- Multiserver Queues: in steady state, the time to empty the system has a moment of order α if the service times admit a moment of order $\alpha - 1$.
- Generalized Jackson Networks: the stationary maximal dater has a moment of order α iff all service times have moments of order $\alpha - 1$.

4 Bounds for Sub-exponential Tail Asymptotics

4.1 Assumptions and Notation

4.1.1 Tails

Let ξ be a non-negative r. v. with distribution function F such that $\mathbf{P}(\xi > x) \equiv 1 - F(x) \equiv \overline{F}(x) > 0$ for all x. Let ξ_1, ξ_2 be independent copies of ξ.

Definition 1 *ξ has a heavy-tailed ditribution (HT), if, for any $c > 0$,*

$$\mathbf{E} \exp(c\xi) \equiv \int_0^\infty \exp(cx) dF(x) = \infty.$$

Definition 2 *ξ has a long-tailed distribution (LT), if, for any $y > 0$,*

$$\overline{F}(x + y) \sim \overline{F}(x) \quad as \quad x \to \infty.$$

Any LT distribution is HT.

Definition 3 *ξ has a sub-exponential distribution (SE), if*

$$\mathbf{P}(\xi_1 + \xi_2 > x) \sim 2\overline{F}(x) \quad as \quad x \to \infty.$$

Any SE distribution is LT.

For basic properties of sub-exponential distributions, see e.g. [9]. In this section, we assume that (IA) and (AA) hold.

Here and later in the paper, for strictly positive functions f and g, the equivalence $f(x) \sim dg(x)$ with $d > 0$ means $f(x)/g(x) \to d$ as $x \to \infty$. This equivalence may also be rewritten as $f(x) = dg(x)(1 + o(1)) = dg(x) + o(g(x)) = dg(x) + o(f(x))$, where $o(1)$ is a function which tends to 0 as x tends to ∞, and $o(g(x))$ is a function such that $o(g(x))/g(x) \to 0$ as $x \to \infty$. By convention, the equivalence $f(x) \sim dg(x)$ with $d = 0$ means $f(x) = o(g(x))$. We will also use the notation $f(x) = \Theta(g(x))$ to mean $\limsup f(x)/g(x) < \infty$ and $\liminf f(x)/g(x) > 0$.

4.1.2 Network Assumptions

We consider a monotone separable network and we assume that there exists a distribution function F on $\mathbb{R}^+$ such that the following holds:
(SE)

(a) F is sub-exponential, with finite first moment

$$M = \int_0^\infty \overline{F}(u) du,$$

where $\overline{F}(u) = 1 - F(u)$ is the tail distribution of F.

(b) The integrated distribution of F:

$$F^s(x) = \frac{1}{M}\int_0^x \overline{F}(u)du$$

is sub-exponential; we will call second tail of F the tail of the integrated distribution, namely $\overline{F}^s(x) = 1 - F^s(x)$.

(c) The following equivalence holds when x tends to ∞:

$$\mathbf{P}(Y_1^{(j)} > x) \sim d^{(j)}\overline{F}(x),$$

for all $j = 1, \ldots, r$ with $\sum_j d^{(j)} \equiv d > 0$.

Under (SE), the following holds:

$$\frac{1}{\mathbf{E}Y_1^{(j)}}\int_x^\infty \mathbf{P}(Y_1^{(j)} > y)dy \sim c^{(j)}\overline{F}^s(x) \tag{14}$$

as $x \to \infty$, where

$$c^{(j)} = \frac{d^{(j)}M}{b^{(j)}}, \quad b^{(j)} = E(Y_1^{(j)}). \tag{15}$$

We also introduce the following assumption:

$$(\mathbf{H}) \quad \mathbf{P}(\sum_1^r Y_1^{(j)} > x) \sim \mathbf{P}(\max_{1\le j\le r} Y_1^{(j)} > x) \sim \sum_1^r \mathbf{P}(Y_1^{(j)} > x) \sim \sum_1^r d^{(j)}\overline{F}(x). \tag{16}$$

Note that the very last equivalence follows from (SE). Assumption (H) is for instance satisfied in the particular case when the random variables $Y_1^{(j)}$ are mutually independent; in Section 4.5 below, we will give sufficient conditions for (H) to hold that go beyond this particular case.

Take any $1 \le i_1, i_2 \le r$, $i_1 \ne i_2$. Since

$$\mathbf{P}(\max_j Y_1^{(j)} > x) \le \sum_j \mathbf{P}(Y_1^{(j)}) - \mathbf{P}(Y_1^{(i_1)} > x, Y_1^{(i_2)} > x),$$

we deduce from (16) that

$$\mathbf{P}(Y_1^{(i_1)} > x, Y_1^{(i_2)} > x) = o(\overline{F}(x)). \tag{17}$$

Remark 1 *In what follows, we will not need i.i.d. assumptions on the interarrival times $\{\tau_n\}$. The results we prove will hold also in the more general situation when these variables satisfy the following three conditions:*

- *$\{\tau_n\}$ forms a stationary ergodic sequence with a finite mean $\mathbf{E}\tau_1 = a$;*
- *$\{\tau_n\}$ is independent of $\{Y_n^{(j)}, j = 1, \ldots, r\}$;*
- *For all $d < a$,*

$$\mathbf{P}\left(\sup_{n\ge 0}\left(nd - \sum_{i=-n}^{-1}\tau_i\right) > x\right) = o(\overline{F}^s(x)).$$

4.2 Veraverbeke's Theorem

We now give a slightly extended version of Veraverbeke's theorem:

Theorem 3 *Let $\{\xi_n\}$ be an i.i.d. sequence with negative mean $\mathbf{E}\xi_1 = -\alpha$, $S_n = \sum_1^n \xi_i$, and $\overline{S} = \sup_n S_n$. Assume that there exists a distribution F on $[0,\infty)$ such that F^s is sub-exponential and $\mathbf{P}(\xi_1 > x) \sim d\overline{F}(x)$ with $d > 0$ as $x \to \infty$. Then, as $x \to \infty$,*

$$\begin{aligned}
\mathbf{P}(\overline{S} > x) &= (1+o(1))\mathbf{P}(\bigcup_n \{\xi_n > x + n\alpha\}) \\
&= (1+o(1))\sum_n \mathbf{P}(\xi_n > x + n\alpha) \\
&= (1+o(1))\frac{1}{\alpha}\int_x^\infty \mathbf{P}(\xi_1 > t)dt \\
&= (1+o(1))\frac{dM}{\alpha}\overline{F}^s(x).
\end{aligned}$$

In particular, consider a GI/GI/1/∞ queue with i.i.d. service times $\{\sigma_n\}$ (with mean b) and i.i.d. interarrival times $\{\tau_n\}$ (with mean $a > b$) and put $\xi_n = \sigma_n - \tau_n$. Assume that $\mathbf{P}(\sigma_1 > x) \sim d\overline{F}(x)$, with F as above. Then the stationary waiting time W is such that

$$\begin{aligned}
\mathbf{P}(W > x) &= (1+o(1))\mathbf{P}(\bigcup_n \{\sigma_n > x + n(a-b)\}) \\
&= (1+o(1))\sum_n \mathbf{P}(\sigma_n > x + n(a-b)) \\
&= (1+o(1))\frac{dM}{a-b}\overline{F}^s(x).
\end{aligned}$$

In particular if the distribution function of σ is F, then

$$\mathbf{P}(W > x) = (1+o(1))\frac{\rho}{1-\rho}\overline{F}^s(x)$$

with $\rho = \frac{b}{a}$.

4.3 Upper Bound

Let $Z = Z_{(-\infty,0]}$ and let L be the integer defined in §2.6 and let $\widehat{s}$ be the service time in the associated L–upper–bound GI/GI/1/∞ queue.

Put $\widehat{b} = \mathbf{E}\widehat{s}$ and note that $\mathbf{E}\widehat{\tau} = La$. Then $\widehat{\rho} = \frac{\widehat{b}}{La} = \lambda\gamma(0)(1+\delta) < 1$ where δ may be chosen as small as possible. We deduce from (7) and (H) that

$$\mathbf{P}(\widehat{s}_1 > x) \sim dL\overline{F}(x).$$

Thus, from Veraverbeke's theorem (more precisely from its extension to response times),

$$\begin{aligned}
\mathbf{P}(\widehat{R} > x) &\sim \frac{1}{La - \widehat{b}}\int_x^\infty \mathbf{P}(\widehat{s} > y)dy \\
&\sim \frac{1}{La - \widehat{b}}\int_x^\infty dL\overline{F}(y)dy \\
&= \frac{\widehat{\rho}Md}{1-\widehat{\rho}}\frac{L}{\widehat{b}}\overline{F}^s(x).
\end{aligned}$$

Here $\frac{L}{\widehat{b}} \to \frac{1}{\gamma(0)}$ as $L \to \infty$. We have proved:

Lemma 9 *Under the (IA), (AA), (SE) and (H) assumptions,*

$$\limsup_{x\to\infty} \frac{\mathbf{P}(Z>x)}{\overline{F}^s(x)} \leq \lim_{x\to\infty} \frac{\mathbf{P}(\widehat{R}>x)}{\overline{F}^s(x)} = \frac{\lambda M d}{1-\lambda\gamma(0)} = \frac{Md}{a-\gamma(0)}. \tag{18}$$

4.4 Lower Bound

From (8),

$$Z = Z_{(-\infty,0]} \quad \geq \quad \underline{R} = \max_j \sup_{n\leq 0} \left(\sum_{i=n}^{0} Y_i^{(j)} - \sum_{i=n}^{-1} \tau_i \right) \equiv \max_j R^{(j)}.$$

Then from Veraverbeke's theorem

$$\mathbf{P}(R^{(j)}>x) \sim \frac{b^{(j)}c^{(j)}}{a-b^{(j)}}\overline{F}^s(x) = \frac{d^{(j)}M}{a-b^{(j)}}\overline{F}^s(x),$$

with $b^{(j)} = \mathbf{E}Y_1^{(j)}$.

Note that

$$\sum_j \mathbf{P}(R^{(j)}>x) \geq \mathbf{P}(\max_j R^{(j)}>x) \geq \sum_j \mathbf{P}(R^{(j)}>x) - \sum_{i_1\neq i_2} \mathbf{P}(R^{(i_1)}>x, R^{(i_2)}>x) \tag{19}$$

and, for $b = \min(b^{(i_1)}, b^{(i_2)})$,

$$\begin{aligned}
\mathbf{P}(R^{(i_1)}>x, R^{(i_2)}>x) &= \sum_{m_1=1}^{\infty}\sum_{m_2=1}^{\infty} \mathbf{P}(Y_{m_1}^{(i_1)} > x+m_1b^{(i_1)}, Y_{m_2}^{(i_2)} > x+m_2b^{(i_2)}) \\
&\leq \sum_{m_1\neq m_2} \mathbf{P}(Y_1^{(i_1)} > x+m_1b^{(i_1)})\mathbf{P}(Y_1^{(i_2)} > x+m_2b^{(i_2)}) \\
&\quad + \sum_{m=1}^{\infty} \mathbf{P}(\max(Y_1^{(i_1)}, Y_1^{(i_2)}) > x+mb) \\
&\leq \sum_{m_1} \mathbf{P}(Y_1^{(i_1)} > x+m_1b^{(i_1)}) \sum_{m_2} \mathbf{P}(Y_1^{(i_2)} > x+m_2b^{(i_2)}) \\
&\quad + \sum_{m=1}^{\infty} o(\overline{F}(x+mb)) \\
&= \Theta((\overline{F}^s(x))^2) + o(\overline{F}^s(x)) = o(\overline{F}^s(x)).
\end{aligned} \tag{20}$$

Thus, we deduce from (19) that under Assumption (H),

$$\mathbf{P}(\underline{R}>x) \sim \overline{F}^s(x) \sum_j \frac{d^{(j)}M}{a-b^{(j)}}. \tag{21}$$

Therefore:

Lemma 10 *Under Assumptions (IA), (AA), (SE) and (H),*

$$\liminf_{x\to\infty} \frac{\mathbf{P}(Z>x)}{\overline{F}^s(x)} \geq \lim_{x\to\infty} \frac{\mathbf{P}(\underline{R}>x)}{\overline{F}^s(x)} = \sum_{j=1}^{r} \frac{d^{(j)}M}{a-b^{(j)}} = \sum_{j=1}^{r} \frac{\rho^{(j)}c^{(j)}}{1-\rho^{(j)}}, \tag{22}$$

with $\rho^{(j)} = \frac{b^{(j)}}{a}$.

Remark 2

- *The asymptotics for the lower and upper bounds are the same up to multiplicative constants.*
- *In the single-server isolated queue case,* $\gamma(0) = b = M$, $\widehat{b} = Lb$ *and* $d = 1$. *Therefore in this case, the upper and lower bounds coincide.*

□

4.5 Relaxing the Independence Assumptions

The aim of this section is to give conditions under which Assumption (H) of §4.1 is satisfied, although the r.v.'s $R^{(j)}$ are not independent.

We assume that there exists a random variable ν taking values in an arbitrary measurable space $(\mathcal{Y}, \mathcal{B}_{\mathcal{Y}})$ and such that

- Given ν, the random variables $Y_1^{(j)}, j = 1, \ldots, r$ are conditionally independent;
- For any $j = 1, \ldots, r$,
$$\mathbf{P}(Y_1^{(j)} > x \mid v) \sim d_\nu^{(j)} \overline{F}(x) \tag{23}$$
$\mathbf{P}_\nu$-a.s., where $d_\nu^{(j)}$ is a non-negative random variable with a finite mean $d^{(j)}$.

Then

$$\tilde{d}_\nu^{(j)} \equiv \sup_x \frac{\mathbf{P}(Y_1^{(j)} > x \mid \nu)}{\overline{F}(x)} \tag{24}$$

is an a.s. finite random variable, too.

Assume in addition that, for any $1 \leq j_1 \leq j_2 \leq r$,

$$\mathbf{E} \prod_{j=j_1}^{j_2} \tilde{d}_\nu^{(j)} < \infty. \tag{25}$$

Lemma 11 *Under the foregoing assumptions, for any* $1 \leq j_1 \leq j_2 \leq r$,

$$\mathbf{P}(\sum_{j=j_1}^{j_2} Y_1^{(j)} > x) \sim \mathbf{P}(\max_{j_1 \leq j \leq j_2} Y_1^{(j)} > x) \sim \sum_{j=j_1}^{j_2} \mathbf{P}(Y_1^{(j)} > x) \sim \sum_{j=j_1}^{j_2} d^{(j)} \overline{F}(x). \tag{26}$$

The proof is given in Appendix 6.2.

Remark 3 *Consider the following example, which covers the Jackson network case. Assume that there are given*

- *Some random vector* $\nu = (\nu^{(1)}, \ldots, \nu^{(r)})$ *with non-negative integer-valued coordinates, such that* $\mathbf{E}\exp(c\nu^{(j)}) < \infty$ *for some* $c > 0$ *and for all* $j = 1, \ldots, r$*;*
- r *sequences* $\{\sigma_n^{(j)}\}$ *of i.i.d. sub-exponential random variables that are mutually independent and do not depend on* ν*, and* $\mathbf{P}(\sigma_1^{(j)} > x) \sim l^{(j)}\overline{F}(x)$*. We do not make the assumption that the r.v.'s* $\nu^{(1)}, \ldots, \nu^{(r)}$ *are independent.*

Put

$$Y_1^{(j)} = \sum_{i=1}^{\nu^{(j)}} \sigma_i^{(j)}.$$

The above conditions imply that, first, for $c^{'} = c/r$*,*

$$\mathbf{E}\exp(c'\sum_j \nu^{(j)}) \le \mathbf{E}\exp(c\max_j \nu^{(j)}) \le \sum_j \mathbf{E}\exp(c\nu^{(j)}) < \infty$$

and, second, for $j = 1, \ldots, r$*,*

$$u^{(j)} \equiv \sup_t \frac{\mathbf{P}(\sigma_1^{(j)} > t)}{\overline{F}(t)} < \infty.$$

Due to sub-exponentiality, for $j = 1, \ldots, r$*,*

$$\mathbf{P}(Y_1^{(j)} > x \mid \nu) \sim \nu^{(j)} l^{(j)} \overline{F}(x).$$

It is known (see, e.g. [9], p.41) that, for any $\varepsilon > 0$*, one can choose* $K^{(j)} \equiv K^{(j)}(\varepsilon)$ *such that*

$$\mathbf{P}(Y_1^{(j)} > x \mid \nu) \le K^{(j)}(1+\varepsilon)^{\nu^{(j)}} \mathbf{P}(\sigma_1^{(j)} > x).$$

The RHS of the latter inequality is not bigger than $K^{(j)}u^{(j)}(1+\varepsilon)^{\nu^{(j)}}\overline{F}(x)$*.*

Take $\varepsilon > 0$ *such that* $\log(1+\varepsilon) \le c^{'}$*. Then the conditions of Lemma 11 are satisfied with* $d_\nu^{(j)} = \nu^{(j)} l^{(j)}$*,* $d^{(j)} = l^{(j)}\mathbf{E}\nu^{(j)}$*, and* $\tilde{d}_\nu^{(j)} = K^{(j)}u^{(j)}(1+\varepsilon)^{\nu^{(j)}}$*.*

4.6 Example: Tandem Queues

The definitions and notation are those of §2.7.1. We assume that

$$\overline{F}_i(x) = \mathbf{P}(\sigma^{(i)} > x) \sim d^{(i)}\overline{F}(x), \tag{27}$$

that $d \equiv d^{(1)} + d^{(2)} > 0$ and that both F and its second tail F^s are sub-exponential. Without loss of generality, one can replace random inter-arrival times by their mean (see §6.1). Assumption (H) is then valid since $\sigma_n^{(1)}$ and $\sigma_n^{(2)}$ are independent.

Denote by Z the stationary sojourn time in the network. We look for the asymptotic behavior of the function $\mathbf{P}(Z > x)$ as $x \to \infty$.

The lower bound (22) is:

$$\liminf_{x\to\infty} \frac{\mathbf{P}(Z > x)}{\overline{F}^s(x)} \quad \ge \quad \frac{\rho^{(1)}c^{(1)}}{1-\rho^{(1)}} + \frac{\rho^{(2)}c^{(2)}}{1-\rho^{(2)}} = \frac{d^{(1)}M}{a-b^{(1)}} + \frac{d^{(2)}M}{a-b^{(2)}}.$$

Since $\gamma(0) = b \equiv \max(b^{(1)}, b^{(2)})$, the upper bound (18) reads:

$$\limsup_{x\to\infty} \frac{\mathbf{P}(Z > x)}{\overline{F}^s(x)} \le \frac{\lambda d M}{1-\lambda b} = \frac{dM}{a-b}.$$

This upper bound was first proved in [6].

5 Exact Tail Asymptotics

5.1 Veraverbeke's Theorem Revisited and How Rare Events Occur

This section contains the proof of a corollary of Veraverbeke's theorem, which provides some indications on how rare events occur in this framework. The notation and certain ideas of this proof will be used in the sharp asymptotics later on.

Consider a GI/GI/1/∞ queue with mean inter-arrival times $a = \mathbf{E}\tau_n$ and mean service times $b = \mathbf{E}\sigma_n$, where $a > b$. Put $\rho = \frac{b}{a} < 1$. Assume the second tail of σ:

$$\overline{F}^s(x) = \frac{1}{b}\int_x^\infty \mathbf{P}(\sigma > u)du$$

to be sub-exponential. Then, in particular,

$$\mathbf{P}(\sigma > x) = o\left(\overline{F}^s(x)\right). \tag{28}$$

As a corollary, there exists a sequence $N \equiv N_x$ such that $N_x \to \infty$ when $x \to \infty$ and such that

$$\sum_{n=0}^{N} \mathbf{P}(\sigma > x + nb) = o\left(\overline{F}^s(x)\right), \quad x \to \infty. \tag{29}$$

Note also that one can choose $z_x \to \infty$, $z_x = o(x)$ such that

$$\frac{\overline{F}^s(x + z_x)}{\overline{F}^s(x)} \to 1 \quad \text{and} \quad \frac{\overline{F}^s(x)}{\overline{F}^s(x - z_x)} \to 1 \tag{30}$$

as $x \to \infty$. For $\xi_n = \sigma_n - \tau_n$,

$$\mathbf{P}(\xi_1 > x) \geq \mathbf{P}(\sigma_1 > x + z_x)\mathbf{P}(\tau_1 \leq z_x) \equiv \mathbf{P}(\sigma_1 > x + z_x)(1 - \Delta_x), \tag{31}$$

where $\Delta_x \to 0$ as $x \to \infty$.

Put $\xi_n = \sigma_n - \tau_n$, $S_{-n} = \sum_{-n}^{-1} \xi_i$, $S_0 = 0$, $S_{-\infty} = -\infty$. Put $S^*_{-n} = \max_{0 \leq j \leq n} S_{-j}$.

Corollary 5 *Let W (resp. R) denote the stationary waiting (resp. response) time in the FIFO GI/GI/1/∞ queue. There exists a sequence ε_n such that $\varepsilon_n \downarrow 0$ and $n\varepsilon_n \uparrow \infty$, when $n \uparrow \infty$ and such that when denoting G_n the interval $G_n = (n(b - a - \varepsilon_n), n(b - a + \varepsilon_n))$ and A_x the event*

$$A_x = \bigcup_{n=N_x}^{\infty} A_{n,x} \tag{32}$$

$$A_{n,x} = \{S^*_{-n} \leq x, S_{-n} \in G_n, \xi_{-n-1} > x + n(a - b + \varepsilon_n)\}, \tag{33}$$

then for all sets A'_x such that

$$\mathbf{P}(A'_x, W > x) \geq \mathbf{P}(A_x, W > x) + o(\overline{F}^s(x)), \tag{34}$$

we have

$$\mathbf{P}(W > x) \sim \mathbf{P}(A'_x \cap \{W > x\}). \tag{35}$$

Similarly, for all sets A'_x such that

$$\mathbf{P}(A'_x, R > x) \geq \mathbf{P}(A_x, R > x) + o(\overline{F}^s(x)) \tag{36}$$

we have

$$\mathbf{P}(R > x) \quad \sim \quad \mathbf{P}(A'_x \cap \{R > x\}), \tag{37}$$

when x tends to ∞.

Note that the event A_x (which will be referred to as the *typical event* in what follows) occurs if there is only one big service time and all other service times or inter-arrival times are approximately equal to their means.

Proof We have $W = \sup_{n\geq 0} S_{-n}$. For all $\varepsilon > 0$ and for all x,

$$\liminf_{n\to\infty} \mathbf{P}\Big(S^*_{-n} \leq x, S_{-n} \in (n(b-a-\varepsilon), n(b-a+\varepsilon))\Big) \geq \mathbf{P}(W < x). \tag{38}$$

Therefore, one can choose a sequence $\varepsilon_n \downarrow 0$, with $n\varepsilon_n \to \infty$ such that

$$\liminf_{n\to\infty} \mathbf{P}(S^*_{-n} \leq x, S_{-n} \in G_n) \geq \mathbf{P}(W < x)$$

for any x. Thus

$$1 - \delta_{n,x} \equiv \inf_{m\geq n} \mathbf{P}(S^*_{-m} \leq x, S_{-m} \in G_m), \tag{39}$$

is such that $\delta_{n,x}$ tends to 0 as $n, x \to \infty$.

We have

$$\begin{aligned}
\mathbf{P}(W > x) &\geq \mathbf{P}(A'_x, W > x) \\
&\geq \mathbf{P}(A_x, W > x) + o(\overline{F}^s(x)) \\
&\geq \mathbf{P}\left(\bigcup_{n=N}^{\infty} \{S^*_{-n} \leq x, S_{-n} \in G_n, \xi_{-n-1} > x + n(a-b+\varepsilon_n)\}\right) + o(\overline{F}^s(x)).
\end{aligned}$$

Since the sets in the last union are disjoint for all sufficiently large x, the probability of the union is the sum of the probabilities. This and the independence assumption imply

$$\mathbf{P}(W > x) \geq \sum_{n=N}^{\infty} \mathbf{P}(S^*_{-n} \leq x, S_{-n} \in G_n)\mathbf{P}(\xi_1 > x + n(a-b+\varepsilon_n)) + o(\overline{F}^s(x)).$$

Using now (39) and (31), we get

$$\begin{aligned}
\mathbf{P}(W > x) &\geq (1-\delta_{N,x}) \sum_{n=N}^{\infty} \mathbf{P}(\xi_1 > x + n(a-b+\varepsilon_n)) + o(\overline{F}^s(x)) \\
&\geq (1-\delta_{N,x})(1-\Delta_x) \sum_{N}^{\infty} \mathbf{P}(\sigma > x + z_x + n(a-b+\varepsilon_n) + o(\overline{F}^s(x))) \\
&\geq (1-\delta_{N,x})(1-\Delta_x) \sum_{0}^{\infty} \mathbf{P}(\sigma > x + z_x + n(a-b+\varepsilon_n))
\end{aligned}$$

$$
\begin{aligned}
& - \sum_{n=0}^{N-1} \mathbf{P}(\sigma > x + z_x + n(a-b+\varepsilon_n) + o(\overline{F}^s(x))) \\
\geq & \frac{(1-\delta_{N,x})(1-\Delta_x)}{a-b+\varepsilon_N} \int_{x+z_x}^{\infty} \mathbf{P}(\sigma > y) dy + o\left(\overline{F}^s(x)\right) \\
= & \frac{(1-\delta_{N,x})(1-\Delta_x)\rho}{1-\rho+\varepsilon_N/a}(1+o(1))\overline{F}^s(x) + o\left(\overline{F}^s(x)\right).
\end{aligned}
$$

So when $x \to \infty$,

$$
P(W > x) \geq P(A'_x, W > x) \geq \frac{\rho}{1-\rho}\overline{F}^s(x)(1+o(1)).
$$

But from Veraverbeke's theorem,

$$
\lim_{x\to\infty} \frac{\mathbf{P}(W > x)}{\overline{F}^s(x)} = \frac{\rho}{1-\rho},
$$

so that

$$
\mathbf{P}(W > x) \quad \geq \quad \mathbf{P}(A'_x, W > x) \geq \frac{\rho}{1-\rho}\overline{F}^s(x)(1+o(1)) = \mathbf{P}(W > x)(1+o(1)).
$$

This concludes the proof of (35).

Using the relation $R = W + \sigma$, the independence and the fact that the tail of W is heavier than that of σ (see (28)), one gets

$$
\lim_{x\to\infty} \frac{\mathbf{P}(R > x)}{\overline{F}^s(x)} = \frac{\rho}{1-\rho}.
$$

The proof of (37) follows from this relation in the same way as in the proof of (35). □

Here are a few examples of sets A'_x:

1. Taking $A'_x = A_x$, we get

$$
\mathbf{P}(W > x) \quad \sim \quad \mathbf{P}(A_x) \sim \sum_{n \geq N_x} \mathbf{P}(A_{n,x} \cap \{W > x\}); \tag{40}
$$

2. Take

$$
A'_x = \bigcup_{n=N_x}^{\infty} \{\sigma_{-n} > x + n(a-b+\varepsilon_n)\}.
$$

We get

$$
\mathbf{P}(W > x) \quad \sim \quad \mathbf{P}\Big(W > x, \bigcup_{n=0}^{\infty} \{\sigma_n > x + n(a-b+\varepsilon_n)\}\Big). \tag{41}
$$

Condition (34) is satisfied since $A_x \subset A'_x$.

3. Take

$$A'_x = \bigcup_{n \geq N_x} A'_{n,x} \tag{42}$$

$$A'_{n,x} = \{S^*_{-n} \leq x, S_{-n} \in G_n, \xi_{-n-1} > x + n(a - b + \Delta_n)\}, \tag{43}$$

where Δ_n is any sequence converging to 0. In order to prove (34) it is enough to show that $\mathbf{P}(A_x \cap (A'_x)^c) = o(\overline{F}^s(x))$, which follows from the bound

$$\begin{aligned}
&\mathbf{P}\Big(\bigcup_{\substack{n \geq N_x \\ \Delta n \geq \varepsilon n}} \{\xi_{-n-1} > x + n(a - b - \varepsilon_n)\} \Big) - \mathbf{P}\Big(\bigcup_{\substack{n \geq N_x \\ \Delta n \geq \varepsilon n}} \{\xi_{-n-1} > x + n(a - b - \Delta_n)\} \Big) \\
\leq &\sum_{\substack{n \geq 1 \\ \Delta_n \geq \varepsilon n}} \mathbf{P}(\xi_{-n-1} \in (x + n(a - b + \varepsilon_n), x + (a - b + \Delta_n)) \\
= &\ o(\overline{F}^s(x)).
\end{aligned}$$

Remark 4 *Let $A'_x = \cup_p A^p_x$. Assume that A'_x satisfies the assumptions of Corollary 5 and the bounds*

$$\begin{aligned}
\mathbf{P}(A'_x, W > x) &\geq \sum_p \mathbf{P}(A^p_x, W > x) + o(\overline{F}^s(x)) \\
\sum_p \mathbf{P}(A^p_x, W > x) &\geq \mathbf{P}(A_x, W > x) + o(\overline{F}^s(x)),
\end{aligned}$$

then we also have

$$\mathbf{P}(W > x) \sim \sum_p \mathbf{P}(A^p_x, W > x). \tag{44}$$

□

We now give a more general version of Corollary 5 which will be useful in what follows. Let p be some discrete index. For all x, n and p, let $I^p_{n,x}$ and $J^p_{n,x}$ be some events. Associated with these events, define

$$\begin{aligned}
C^p_{n,x} &= \{S^*_{-n} \leq x, S_{-n} \in G_n, J^p_{n,x}\}, \\
B^p_{n,x} &= C^p_{n,x} \bigcap I^p_{n,x} \\
B^p_x &= \bigcup_{n=N_x} B^p_{n,x}
\end{aligned}$$

Corollary 6 *Assume the events $I^p_{n,x}$ and $J^p_{n,x}$ satisfy the following assumptions:*

- *for all n, p and x, the events $C^p_{n,x}$ and $I^p_{n,x}$ are independent;*
- *the function $\delta_{n,x}$ defined by:*

$$1 - \delta_{n,x} \equiv \inf_{m \geq n} \inf_p \mathbf{P}(C^p_{n,x})$$

is such that $\delta_{n,x} \to 0$ as $n, x \to \infty$;

- *the events $I^p_{n,x}$ are such that:*
  1. *for all n, on the event $C^p_{n,x}$,*
  $$I^p_{n,x} \subset \{S_{-n-1} > x\};$$
  2. *for all n, the events $C^p_{n,x} \bigcap I^p_{n,x}$ are disjoint in p for sufficiently large x;*
  3. *when $x \to \infty$,*
  $$\sum_{n=N_x}^{\infty} \sum_p \mathbf{P}(I^p_{n,x}) = \frac{\rho}{1-\rho} \overline{F}^s(x) + o(\overline{F}^s(x)).$$

Then for all events A'_x such that

$$\mathbf{P}(A'_x, W > x) \geq \mathbf{P}\left(\bigcup_p B^p_x, W > x\right) + o(\overline{F}^s(x)),$$

we have

$$\mathbf{P}(W > x) \quad \sim \quad \mathbf{P}(W > x, A'_x). \tag{45}$$

Similarly, for all events A'_x such that

$$\mathbf{P}(A'_x, R > x) \geq \mathbf{P}\left(\bigcup_p B^p_x, R > x\right) + o(\overline{F}^s(x)),$$

we have

$$\mathbf{P}(R > x) \quad \sim \quad \mathbf{P}(R > x, A'_x). \tag{46}$$

Proof The arguments are similar to those in the proof of Corollary 5. Assumption (1) implies that the events $\bigcup_p B^p_{n,x}$ are disjoint in n. In view of Assumption (2), the sets $B^p_{n,x}$ are actually disjoint in n and p, at least for x large, so that

$$\begin{aligned}
\mathbf{P}(W > x) &\geq \mathbf{P}(W > x, A'_x) \\
&\geq \mathbf{P}\left(W > x, \bigcup_{n=N}^{\infty} \bigcup_p B^p_{n,x}\right) + o(\overline{F}^s(x)) \\
&= \sum_{n=N}^{\infty} \sum_p \mathbf{P}(B^p_{n,x}) + o(\overline{F}^s(x)).
\end{aligned}$$

From this and the independence assumptions, we get

$$\begin{aligned}
\mathbf{P}(W > x) &\geq \mathbf{P}(W > x, A'_x) + o(\overline{F}^s(x)) \\
&\geq \sum_{n=N}^{\infty} \sum_p \mathbf{P}(C^p_{n,x}) \mathbf{P}(I^p_{n,x}) + o(\overline{F}^s(x)) \\
&\geq (1 - \delta_{N,x}) \sum_{n=N}^{\infty} \sum_p \mathbf{P}(I^p_{n,x}) + o(\overline{F}^s(x)).
\end{aligned}$$

Using Assumption (3), we get

$$\mathbf{P}(W > x) \quad \geq \quad \mathbf{P}(W > x, A'_x) \geq \frac{\rho}{1-\rho} \overline{F}^s(x)(1 + o(1)),$$

and the proof of (45) follows from Veraverbeke's theorem in a similar way. □

5.2 The Key Equivalences for Exact Asymptotics

Theorem 4 *Let Z be the stationary maximal dater of some monotone separable network satisfying (IA), (AA), (H) and (SE). Then denoting $\widehat{A}_x$ the typical event of the L–upper–bound queue*[1]*, we have:*

$$\mathbf{P}(Z > x) \sim \mathbf{P}(Z > x, \widehat{A}_x). \tag{47}$$

In addition,

$$\mathbf{P}(Z > x, \hat{A}_x) = \Theta\left(\overline{F}^s(x)\right). \tag{48}$$

The same results hold true for any set $\widehat{A}'_x$ satisfying the properties listed in Corollary 6 w.r.t. $\widehat{A}_x$.

Proof It is enough to prove the property for $\widehat{A}'_x$ as above. We obviously have $\mathbf{P}(Z > x) \geq \mathbf{P}(Z > x, \widehat{A}'_x)$. So, it is enough to show that $\mathbf{P}(Z > x) \leq (1 + o(1))\mathbf{P}(Z > x, \widehat{A}'_x)$. For this, we use the following inequalities:

$$\begin{aligned} \mathbf{P}(Z > x) &= \mathbf{P}(Z > x, \widehat{A}'_x) + \mathbf{P}(Z > x, \widehat{R} > x, (\widehat{A}'_x)^c) \\ &\leq \mathbf{P}(Z > x, \widehat{A}'_x) + \mathbf{P}(\widehat{R} > x, (\widehat{A}'_x)^c), \end{aligned}$$

where A^c denotes the complement of set A. From (46),

$$\mathbf{P}(\widehat{R} > x, (\widehat{A}'_x)^c) = o\left(\mathbf{P}(\widehat{R} > x)\right).$$

Using now the fact that

$$\liminf_{x\to\infty} \frac{\mathbf{P}(\underline{R} > x)}{\mathbf{P}(\widehat{R} > x)} > 0, \tag{49}$$

which directly follows from Lemmas 9 and 10, we obtain:

$$\mathbf{P}(\widehat{R} > x, (\widehat{A}'_x)^c) = o\left(\mathbf{P}(\widehat{R} > x)\right) = o\left(\mathbf{P}(\underline{R} > x)\right).$$

Finally since $\underline{R} \leq Z$,

$$\mathbf{P}(\widehat{R} > x, (\widehat{A}'_x)^c) = o\left(\mathbf{P}(\widehat{R} > x)\right) = o\left(\mathbf{P}(\underline{R} > x)\right) = o\left(\mathbf{P}(Z > x)\right).$$

So

$$\mathbf{P}(Z > x) \leq \mathbf{P}(Z > x, \widehat{A}'_x) + o\left(\mathbf{P}(Z > x)\right),$$

which concludes the proof of (47) and its extension to $\widehat{A}'_x$.

For (48), the proof of the lim inf property follows from what precedes. That of the lim sup property follows from the fact that $\mathbf{P}(Z > x, \hat{A}_x) \leq \mathbf{P}(\hat{A}_x) = \Theta(\overline{F}^s(x))$. □

Remark 5 *Using the same arguments as above, it is easy to check that for monotone separable networks which satisfy only Assumption (IA), the conclusions of the preceding lemma hold provided the following additional assumptions are satisfied:*

1. *The L–upper–bound queue has service times with sub-exponential first and second tails.*

[1] More generally we will add a hat to indicate that a variable pertains to the upper bound queue

2. There exists a lower bound event $C_x \subset \{Z > x\}$ such that

$$\liminf_{x\to\infty} \frac{\mathbf{P}(C_x)}{\mathbf{P}(\widehat{R} > x)} > 0. \tag{50}$$

□

Let η_n be a real valued sequence converging to 0 and such that, for any $j = 1, \ldots, r$,

$$\mathbf{P}\left(\left|\frac{Y_{-l}^{(j)} + \ldots + Y_{-1}^{(j)}}{l} - b^{(j)}\right| \leq \eta_j, \forall l \geq m\right) \tag{51}$$

tends to 1 as $m \to \infty$. Denote

$$H_{x,n}^{(j)} = \left\{\left|\frac{Y_{-l}^{(j)} + \cdots + Y_{-1}^{(j)}}{l} - b^{(j)}\right| \leq \eta_l, \forall N_x \leq l < n\right\}, \tag{52}$$

$$K_{x,n}^{(j)} = \bigcap_{i\neq j} H_{x,n}^{(i)} \bigcap H_{x,n-1}^{(j)}. \tag{53}$$

Then

$$\max_j \sup_{n\geq N_x} \mathbf{P}\left((K_{x,n}^{(j)})^c\right) \to 0 \tag{54}$$

as $x \to \infty$. Associated with these sets, define

$$\begin{aligned} A_{x,n}^{(j)} &= \{Y_n^{(j)} > x + n(a - \frac{\widehat{b}}{L} + \eta_n)\} \bigcap K_{x,n}^{(j)} \\ A_x^{(j)} &= \bigcup_{n\geq N_x} A_{x,n}^{(j)} \\ \mathcal{A}_x &= \bigcup_{j=1}^{r} A_x^{(j)}. \end{aligned} \tag{55}$$

Theorem 5 *Under the assumptions of Theorem 4,*

$$\mathbf{P}(Z > x) \sim \mathbf{P}(Z > x, \mathcal{A}_x) \tag{56}$$

$$\sim \sum_{j=1}^{r} \mathbf{P}(Z > x, A_x^{(j)}) \tag{57}$$

as $x \to \infty$, for some sequence η_n satisfying (51) and such that $\eta_n \downarrow 0$ when $n \uparrow \infty$. In addition,

$$\mathbf{P}(Z > x) = \mathbf{P}(Z > x, \mathcal{A}_x) + o(\overline{F}^s(x)) \tag{58}$$

$$= \sum_{j=1}^{r} \mathbf{P}(Z > x, A_x^{(j)}) + o(\overline{F}^s(x)) \tag{59}$$

Proof Let us first show that the set $\mathcal{A}_x$ satisfies the properties of A'_x of Corollary 6. Take index p of the form (l, j), with $l \in \{0, \ldots, L-1\}$ and $j \in \{1, \ldots, r\}$ and the following sets:

$$\begin{aligned} I_{n,x}^{l,j} &= \{Y_{-((n+1)L+l)}^{(j)} - \widehat{\tau}_n > x + (n+2)L(a - \frac{\widehat{b}}{L} - \frac{\widehat{\varepsilon}_n}{L})\}, \\ J_{n,x}^{l,j} &= H_{x,(n+1)L+l-1}^{(j)} \bigcap_{i\neq j} H_{x,(n+1)L+l}^{(i)}, \end{aligned}$$

where $\widehat{\varepsilon}_n$ is the sequence associated with the L upper-bound queue.

- for all l and j, and all n and x, the events $I_{n,x}^{l,j}$ and
$$C_{n,x}^{l,j} = \{\widehat{S}_{-n}^{*} \leq x, \widehat{S}_{-n} \in \widehat{G}_n, J_{n,x}^{l,j}\}$$
are independent indeed;
- The corresponding $\delta_{n,x}$ function tends to zero as a consequence of the SLLN (and the fact that $\widehat{R}$ is finite);
- The fact that the set $I_{n,x}^{l,j}$ satisfies Property (1) of Corollary 6 follows from (7), which implies that
$$\widehat{s}_{-n-1} \geq \max_{l=0,\ldots,L-1} \max_{j=1,\ldots,r} Y_{-((n+1)L+l)}^{(j)}.$$
- Property (2) is satisfied for sufficiently large x;
- Property (3) follows directly from (7), (SE), and (H), which imply that
$$\mathbf{P}(\widehat{s}_n > x) \sim \sum_{l=0}^{L-1} \sum_{j=1}^{r} \mathbf{P}(Y_l^{(j)} > x) \sim dL\overline{F}(x).$$

In addition, we have $\mathcal{A}_x \supset \bigcup_p B_x^p$; this follows from the fact that for all l, n, j and x, if one takes $m = (n+1)L + l$, then

$$I_{n,x}^{l,j} \subset \{Y_{-m}^{(j)} > x + m(a - \frac{\widehat{b}}{L} + \eta_m)\},$$

where $\eta_m = \widehat{\varepsilon}_n = \widehat{\varepsilon}_{\lfloor \frac{m}{L} \rfloor - 1}$.

So, from Corollary 6, $\mathbf{P}(\widehat{W} > x) \sim \mathbf{P}(\widehat{W} > x, \mathcal{A}_x)$. The proof of (56) then directly follows from Theorem 4.

Denote $f(x) = \mathbf{P}(Z > x)$, $g(x) = \mathbf{P}(Z_x, \mathcal{A}_x)$. Since $f(x) \sim g(x)$ and $f(x) = \Theta(\overline{F}^s(x))$ (see (48)), then $g(x) = \Theta(\overline{F}^s(x))$ and $f(x) = g(x) + o(\overline{F}^s(x))$, which is (58).

We now prove (57) and (59). We have

$$\mathbf{P}(Z > x, \mathcal{A}_x) \leq \sum_j \mathbf{P}(Z > x, A_x^{(j)}) \tag{60}$$

$$\mathbf{P}(Z > x, \mathcal{A}_x) \geq \sum_j \mathbf{P}(Z > x, A_x^{(j)}) - \sum_{i_1 \neq i_2} \mathbf{P}(A_x^{(i_1)} \bigcap A_x^{(i_2)}). \tag{61}$$

In addition, for all $i_1 \neq i_2$,

$$\mathbf{P}(A_x^{(i_1)} \bigcap A_x^{(i_2)}) \leq \mathbf{P}\left(\bigcap_{r=1}^{2} \bigcup_{n_r=1}^{\infty} \{Y_n^{(i_r)} > x + n(a - \frac{\widehat{b}}{L} + \eta_n)\}\right) \leq \mathbf{P}(R^{(i_1)} > x, R^{(i_2)} > x).$$

The RHS of the last expression is $o(\overline{F}^s(x))$ (see (20)).

Take f and g as above and $h(x) = \sum \mathbf{P}(Z > x, A_x^{(j)})$. From (60)-(61), we get: $g(x) \leq h(x)$ and $g(x) \geq h(x) + o(\overline{F}^s(x))$, from which we deduce that $h(x) = \Theta(\overline{F}^s(x))$ and $g(x) \sim h(x)$. □

The equivalences (47) and (56)-(57) will be the key relationships for the exact asymptotics of the examples of the forthcoming sections. They again indicate that, both for the upper bound queue and for the monotone separable network itself, at most one of the service times is large whereas all other ones are moderate.

5.3 Examples

5.3.1 Tandem Queues – End to End Delay

In this section, we prove the following exact asymptotic, which refines the bounds of § 4.6 (these bounds do not coincide in general):

Theorem 6 *Under the assumptions of § 4.6,*

$$\mathbf{P}(Z > x) \sim \left(\frac{d^{(1)}M}{a-b} + \frac{d^{(2)}M}{a-b^{(2)}} \right) \overline{F}^s(x) \sim \frac{\rho^{(1)}}{1-\rho} \overline{F}_1^s(x) + \frac{\rho^{(2)}}{1-\rho^{(2)}} \overline{F}_2^s(x) \qquad (62)$$

where $b = \max(b^{(1)}, b^{(2)})$, $\rho = \lambda b$ *and* $F_i(x) = P(\sigma^{(i)} \leq x)$.

Remark 6 *As a corollary of Theorem 6 and of results from [11], one can easily derive sharp asymptotics for the stationary queue length* $Q = Q_1 + Q_2$ *in the tandem queue.* □

Proof W.l.o.g we can assume interarrival times to be constants and equal to a (see §6.1 in the Appendix). Note that $Y_n^{(j)} = \sigma_n^{(j)}$, so that (H) trivially holds if service times are independent.

Choose a sequence $N \equiv N_x$ tending to ∞ with x and such that

$$N_x \overline{F}(x) = o\left(\overline{F}^s(x)\right).$$

Then $N_x \mathbf{P}(\sigma_1^{(i)} > x) = o\left(\overline{F}^s(x)\right)$ for $i = 1, 2$.

From (10),

$$Z = \sigma_0^{(2)} + \sup_{-\infty < p \leq q \leq 0} \left(\sum_{m=p}^{q} \sigma_m^{(1)} + \sum_{m=q}^{-1} \sigma_m^{(2)} + pa \right) \equiv \sigma_0^{(2)} + \tilde{Z}, \qquad (63)$$

where $\sum_{m=0}^{-1} = 0$ by convention.

Since $\sigma_0^{(2)}$ and $\tilde{Z}$ are independent and $\mathbf{P}(\sigma_0^{(2)} > x) = \Theta(\overline{F}(x)) = o(\overline{F}^s(x))$, it suffices to show that the distribution of $\tilde{Z}$ has the desired tail.

Lower bound. From Theorem 5,

$$\mathbf{P}(Z > x) = \sum_{j=1}^{2} \mathbf{P}(Z > x, A_x^{(j)}) + o(\overline{F}^s(x)) \geq \sum_{j=1}^{2} \mathbf{P}(\tilde{Z} > x, A_x^{(j)}) + o(\overline{F}^s(x)).$$

For all sufficiently large x, for any $n_1, n_2 \geq N_x$ and $j_1, j_2 \in \{1, 2\}$ such that $(n_1, j_1) \neq (n_2, j_2)$, the events $A_{x,n_1}^{(j_1)}$ and $A_{x,n_2}^{(j_2)}$ do not intersect. Therefore,

$$\sum_{j=1}^{2} \mathbf{P}(\tilde{Z} > x, A_x^{(j)}) = \sum_{n=N_x}^{\infty} \sum_{j=1}^{2} \mathbf{P}(\tilde{Z} > x, A_{x,n}^{(j)}).$$

For $j = 2$,

$$\mathbf{P}(\tilde{Z} > x, A_{x,n}^{(2)}) \geq \mathbf{P}(\sigma_{-n}^{(1)} + \sum_{m=-n}^{-1} \sigma_m^{(2)} - na > x, A_{x,n}^{(2)})$$

$$\begin{aligned}
&\geq \mathbf{P}(\sum_{m=-n}^{-1} \sigma_m^{(2)} - na > x, \sigma_{-n}^{(2)} > x + n(a - \frac{\hat{b}}{L} + \eta_n), K_{x,n}^{(2)}) \\
&\geq \mathbf{P}(\sigma_{-n}^{(2)} + (n-1)(b^{(2)} - \eta_n) - na > x, \sigma_{-n}^{(2)} > x + n(a - \frac{\hat{b}}{L} + \eta_n))\mathbf{P}(K_{x,n}^{(2)}) \\
&= (1 + o(1))\mathbf{P}(\sigma_{-n}^{(2)} > n(a - \min(\frac{\hat{b}}{L}, b^{(2)}) + \eta_n) + b^{(2)}).
\end{aligned}$$

For any $\varepsilon > 0$, one can choose L sufficiently large such that

$$n(a - \min(\frac{\hat{b}}{L}, b^{(2)}) + \eta_n) + b^{(2)} \leq n(a - b^{(2)} + \varepsilon)$$

for all $n \geq N_x$. Therefore,

$$\begin{aligned}
\sum_{n \geq N_x} \mathbf{P}(\tilde{Z} > x, A_{x,n}^{(2)}) &\geq (1 + o(1)) \sum_{n \geq N_x} \mathbf{P}(\sigma_0^{(2)} > x + n(a - b^{(2)} + \varepsilon)) \\
&= (1 + o(1)) \frac{d^{(2)}}{a - b^{(2)} + \varepsilon} \int_x^\infty \overline{F}(y) dy.
\end{aligned}$$

For $j = 1$,

$$\begin{aligned}
\mathbf{P}(\tilde{Z} > x, A_{x,n}^{(1)}) &\geq \mathbf{P}(\sigma_{-n}^{(1)} + \max(\sum_{m=-n+1}^{-1} \sigma_m^{(1)}, \sum_{m=-n}^{-1} \sigma_m^{(2)}) - na > x, A_{x,n}^{(1)}) \\
&\geq \mathbf{P}(\sigma_{-n}^{(1)} + \max((n-1)(b^{(1)} - \eta_n), n(b^{(2)} - \eta_n)) - na > x, \\
&\quad \sigma_{-n}^{(1)} > x + n(a - \frac{\hat{b}}{L} + \eta_n))\mathbf{P}(K_{x,n}^{(1)}) \\
&\geq (1 + o(1))\mathbf{P}(\sigma_{-n}^{(1)} > x + n(a - \min(b, \frac{\hat{b}}{L}) + \eta_n) + b^{(1)}).
\end{aligned}$$

For any $\varepsilon > 0$, one can choose L sufficiently large such that

$$n(a - \min(b, \frac{\hat{b}}{L}) + \eta_n) + b^{(1)} \leq n(a - b + \varepsilon)$$

for all $n \geq N_x$. Therefore,

$$\begin{aligned}
\sum_{n \geq N_x} \mathbf{P}(\tilde{Z} > x, A_{x,n}^{(1)}) &\geq (1 + o(1)) \sum_{n \geq N_x} \mathbf{P}(\sigma_0^{(1)} > x + n(a - b + \varepsilon)) \\
&= (1 + o(1)) \frac{c^{(1)}}{a - b + \varepsilon} \int_x^\infty \overline{F}(y) dy.
\end{aligned}$$

Summing up the bounds in the cases $j = 1$ and $j = 2$, and tending ε to zero, we obtain that the RHS of (62) gives the asymptotic lower bound for the probability $\mathbf{P}(Z > x)$.

We now concentrate on the **upper bound**. Since (AA), (IA), (SE) and (H) are all satisfied, we can use Theorem 5. As $Y_n^{(i)} = \sigma_n^{(i)}$, $i = 1, 2$, the events defined in Theorem 5 take the following special form:

$$H_{x,m}^{(i)} = \{\left| \frac{\sigma_{-l}^{(i)} + \ldots + \sigma_{-1}^{(i)}}{l} - b^{(i)} \right| \leq \eta_l, \forall N \leq l \leq m\}$$

and we have

$$\begin{aligned}\mathbf{P}(Z>x) &\sim \mathbf{P}(Z>x,\mathcal{A}_x)\\ &\le \sum_{n\ge N}\mathbf{P}\left(Z>x,\sigma_{-n}^{(1)}>x+n(a-\widehat{b}/L-\eta_n),H_{x,n-1}^{(1)},H_{x,n}^{(2)}\right)\\ &+ \sum_{n\ge N}\mathbf{P}\left(Z>x,\sigma_{-n}^{(2)}>x+n(a-\widehat{b}/L-\eta_n),H_{x,n}^{(1)},H_{x,n-1}^{(2)}\right)+o(\overline{F}^s(x)).\end{aligned} \tag{64}$$

Let us first concentrate on the first sum of (64). From the bound

$$\begin{aligned}&\mathbf{P}\left(Z>x,\sigma_{-n}^{(1)}>x+n(a-\widehat{b}/L-\eta_n),H_{x,n-1}^{(1)},H_{x,n}^{(2)}\right)\\ \le\ &\mathbf{P}\left(\sigma_{-n}^{(1)}>x+n(a-\widehat{b}/L-\eta_n),H_{x,n-1}^{(1)},H_{x,n}^{(2)}\right),\end{aligned}$$

we obtain via the same technique as above that this first sum is bounded from above by

$$(1+o(1))\left(\frac{d^{(1)}}{a-\widehat{b}/L}\int_x^\infty\overline{F}(y)dy\right),$$

where $\widehat{b}/L$ is as close as one wants to b.

For the second term of (64), we have to consider two cases: if $b^{(2)}\ge b^{(1)}$, then we use the same arguments as above to show that the corresponding sum can be evaluated as

$$(1+o(1))\left(\frac{d^{(2)}}{a-\widehat{b}/L}\int_x^\infty\overline{F}(y)dy\right),$$

where $\widehat{b}/L$ is as close as one wants to $b^{(2)}$.

Consider now the case $b^{(2)}<b^{(1)}$. From (9), we see that Z is the supremum of a denumerable set of terms. From (63), $\widetilde{Z}=\max(X_n,Y_n)$, where X_n is the supremum of the set of terms which contain the variable $\sigma_n^{(2)}$:

$$X_n = \sup_{p\le -n}\ \sup_{p\le q\le -n}\left(\sum_{m=p}^{q}\sigma_m^{(1)}+\sum_{m=q}^{-1}\sigma_m^{(2)}+pa\right)$$

whereas Y_n is the supremum of the set of terms which don't, that is $Y_n=\max(Y_n^1,Y_n^2)$ with

$$\begin{aligned}Y_n^1 &= \sup_{p>-n}\ \sup_{p\le q\le 0}\left(\sum_{m=p}^{q}\sigma_m^{(1)}+\sum_{m=q}^{-1}\sigma_m^{(2)}+pa\right)\\ Y_n^2 &= \sup_{p\le -n}\ \sup_{-n<q\le 0}\left(\sum_{m=p}^{q}\sigma_m^{(1)}+\sum_{m=q}^{-1}\sigma_m^{(2)}+pa\right).\end{aligned}$$

We can rewrite X_n as

$$X_n=\sigma_{-n}^{(2)}+\sum_{m=-n+1}^{-1}\sigma_m^{(2)}-na+V_n,$$

with

$$V_n = \sup_{p\le -n} \sup_{p\le q\le -n} \left(\sum_{m=p}^{q} \sigma_m^{(1)} + \sum_{m=q}^{-n-1} \sigma_m^{(2)} - (-p-n)a \right).$$

On $H_{x,n-1}^{(2)}$, we have

$$X_n \le \sigma_{-n}^{(2)} + (n-1)(b^{(2)} - a + \eta_{n-1}) - a + V_n.$$

So

$$\begin{aligned}
& \mathbf{P}\left(\widetilde{Z} > x, \sigma_{-n}^{(2)} > x + n(a - \widehat{b}/L - \eta_n), H_{x,n}^{(1)}, H_{x,n-1}^{(2)}\right) \\
\le\ & \mathbf{P}\left(X_n > x, \sigma_{-n}^{(2)} > x + n(a - \widehat{b}/L - \eta_n), H_{x,n}^{(1)}, H_{x,n-1}^{(2)}\right) \\
& + \mathbf{P}\left(Y_n > x, \sigma_{-n}^{(2)} > x + n(a - \widehat{b}/L - \eta_n), H_{x,n}^{(1)}, H_{x,n-1}^{(2)}\right) \\
\le\ & \mathbf{P}\left(\sigma_{-n}^{(2)} + (n-1)(b^{(2)} - a + \eta_{n-1}) - a + V_n > x, \sigma_{-n}^{(2)} > x + n(a - \hat{b}/L - \eta_n)\right) \\
& + \mathbf{P}\left(Y_n > x, \sigma_{-n}^{(2)} > x + n(a - \hat{b}/L - \eta_n\right). \qquad (65)
\end{aligned}$$

Let u_n denote the first term of (65) and v_n the second one.

For z_x satisfying (30), there exists a function δ_x' going to 0 as x tends to ∞ and such that $\mathbf{P}(V_n - a < z_x) \ge 1 - \delta_x'$ for all $n \ge N_x$. Indeed,

$$V_n \le_{st} \hat{R} < \infty \quad \text{a.s.}$$

Therefore (note that V_n and $\sigma_{-n}^{(2)}$ are independent),

$$u_n \le (1-\delta_x')\mathbf{P}(\sigma_{-n}^{(2)} > x - z_x + (n-1)(a - b^{(2)} - \eta_{n-1})) + \delta_x' \mathbf{P}(\sigma_{-n}^{(2)} > x + n(a - \hat{b}/L - \eta_n)).$$

Thus,

$$\begin{aligned}
\sum_{n\ge N_x} u_n &\le (1+o(1)) \left(\frac{d^{(2)}}{a - b^{(2)} - \eta_{N-1}} \int_x^\infty \overline{F}(y)dy \right) + o(1) \cdot \left(\frac{d^{(2)}}{a - \hat{b}/L - \eta_N} \int_x^\infty \overline{F}(y)dy \right) \\
&= (1+o(1)) \left(\frac{d^{(2)}}{a - b^{(2)}} \int_x^\infty \overline{F}(y)dy \right). \qquad (66)
\end{aligned}$$

As for the second term v_n of (65), we use the fact that Y_n converges monotonically to some a.s. finite r.v. (the finiteness follows from the fact that for all n, $Y_n \le Z$) and the independence between Y_n and $\sigma_n^{(2)}$ to show that

$$\sum_{n\ge N} v_n \le \mathbf{P}(Z > x) \sum_{n\ge N_x} \mathbf{P}\left(\sigma_{-n}^{(2)} > x + n(a - \widehat{b}/L)\right) = o(\overline{F}^s(x)). \qquad (67)$$

Therefore, in the case $b^{(2)} < b^{(1)}$, we deduce from (65), (66) and (67) that the second sum in (64) is also bounded from above by

$$(1+o(1)) \left(\frac{d^{(2)}}{a - b^{(2)}} \int_x^\infty \overline{F}(y)dy \right)$$

as x tends to ∞, and this concludes the proof of the upper bound. □

5.3.2 Tandem Queues – Delay at the Second Queue

In this section, we focus on the asymptotics for the stationary waiting time $W^{(2)} \equiv W_0^{(2)}$ of customer 0 at the second queue. Results on the matter were obtained by Huang and Sigman in [12]; the case considered in [12] is that where the tail of $\sigma^{(2)}$ is heavier than that of $\sigma^{(1)}$. The results of the present section are more general in that such an assumption is not required. We only assume here that $d^{(2)} > 0$.

First, from $W^{(2)} \leq Z$, we get the following upper bound from (62)

$$\mathbf{P}(W^{(2)} > x) \quad \leq \quad UB(x) \equiv \left(\frac{d^{(1)} M}{a - b} + \frac{d^{(2)} M}{a - b^{(2)}} \right) \overline{F}^s(x) \tag{68}$$

The lower bound

$$\mathbf{P}(W^{(2)} > x) \quad \geq \quad LB(x) \equiv \left(\frac{d^{(2)} M}{a - b^{(2)}} \right) \overline{F}^s(x), \tag{69}$$

also holds (in the proof of the lower bound of Veraverbeke' theorem, we do not need i.i.d. assumptions on interarrival times, but only the SLLN).

Corollary 7 *Under the foregoing assumptions,*

$$\mathbf{P}(W^{(2)} > x) \sim \mathbf{P}(W^{(2)} > x, \widehat{A}_x), \tag{70}$$

where $\widehat{A}_x$ is the typical event of Theorem 4; in addition

$$\mathbf{P}(W^{(2)} > x) \quad \sim \quad \mathbf{P}(W^{(2)} > x, \mathcal{A}_x) \tag{71}$$

$$\mathbf{P}(W^{(2)} > x) \quad \sim \quad \mathbf{P}(W^{(2)} > x, A_x^{(2)}) + \mathbf{P}(W^{(2)} > x, A_x^{(1)}), \tag{72}$$

with $\mathcal{A}_x$, $A_x^{(i)}$ the events of Theorem 5.

Proof The proof of (70) is similar to that of Theorem 4: we also have $W^{(2)} \leq \overline{R}$ and in place of (49), we have (directly from (69))

$$\liminf_{x \to \infty} \frac{\mathbf{P}(W^{(2)} > x)}{\mathbf{P}(\widehat{R} > x)} > 0. \tag{73}$$

The proof of (71) is similar to that of Theorem 5 and will be omitted. □

Case $b^{(1)} > b^{(2)}$ For the following theorem, we do not need any assumption on the tail of $\sigma^{(1)}$. So we only assume that F_2 satisfies the hypothesis (27). In fact, for the proof of the next theorem we even do not need to assume that F is sub-exponential, the only required assumption being that F^s is sub-exponential.

Theorem 7 *If $a > b^{(1)} > b^{(2)}$, then, as $x \to \infty$,*

$$\mathbf{P}(W^{(2)} > x) \sim \frac{d^{(2)} M}{a - b^{(2)}} \overline{F}^s(x). \tag{74}$$

Proof We already established the right lower bound in (69). Thus, it is enough to derive an upper bound which coincides with the lower one.

Let $\tau_i^{(2)}$ denote the inter-arrival times (between i-th and $(i+1)$-st arrivals) in the second station in the stationary regime. Clearly, $\tau_i^{(2)} \geq \sigma_{i+1}^{(1)}$ a.s. Set

$$\xi_i = \sigma_i^{(2)} - \tau_i^{(2)} \quad \text{and} \quad \tilde{\xi}_i = \sigma_i^{(2)} - \sigma_{i+1}^{(1)}.$$

We will use the following property, which follows from the fact that F^s is sub-exponential:

$$\int_x^\infty \mathbf{P}(\xi > t)dt \quad \sim \quad \int_x^\infty \mathbf{P}(\sigma^{(2)} > t)dt \sim d^{(2)} M\overline{F}^s(x) \tag{75}$$

as $x \to \infty$ (see the discussion at the beginning of §5.1).

Put also

$$S_n = \sum_{i=-n}^{-1} \xi_i \quad \text{and} \quad \tilde{S}_n = \sum_{i=-n}^{-1} \tilde{\xi}_i,$$
$$\overline{S} = \sup_n S_n^+ \quad \text{and} \quad \tilde{\overline{S}} = \sup_n \tilde{S}_n^+.$$

Clearly, $W^{(2)} = \overline{S} \leq \tilde{\overline{S}}$ a.s.

Set $c = b^{(2)} - a$ and $\tilde{c} = b^{(2)} - b^{(1)}$. For all $\varepsilon > 0$, $R > 0$, and n, define the event:

$$\begin{aligned} D_{n,\varepsilon,R} \quad = \quad & \Big\{ S_i \leq R - i(c-\varepsilon), \tilde{S}_i \leq R - i(\tilde{c}-\varepsilon),\ i = 1,2,\ldots,n-1; \\ & \tilde{S}_{n+j} - \tilde{S}_n \leq R - j(\tilde{c}-\varepsilon),\ j = 1,2,\ldots \Big\}. \end{aligned}$$

By the SLLN, for any $\varepsilon > 0$, there exists $R > 0$ such that, for any $n = 1,2,\ldots$, $\mathbf{P}(D_{n,\varepsilon,R}) \geq 1 - \varepsilon$. We have

$$\begin{aligned} \mathbf{P}(\overline{S} > x) \quad &= \quad \mathbf{P}(\overline{S} > x, \tilde{\overline{S}} > x) \\ &\leq \quad \sum_n \mathbf{P}(\overline{S} > x, \tilde{\xi}_n > x + n\tilde{c}) + o(\overline{F}^s(x)) \qquad (76) \\ &\leq \quad \sum_n \mathbf{P}(\overline{D}_{n,\varepsilon,R}, \tilde{\xi}_n > x + n\tilde{c}) + \sum_n \mathbf{P}(D_{n,\varepsilon,R}, \overline{S} > x) + o(\overline{F}^s(x)) \\ &\equiv \quad \Sigma_1 + \Sigma_2 + o(\overline{F}^s(x)). \qquad (77) \end{aligned}$$

To show (76), denote $A = \{\overline{S} > x\}$, $B = \{\tilde{\overline{S}} > x\}$, $C = \bigcup_n \{\tilde{\xi}_n > x + n\tilde{c}\}$ and $C_n = \{\tilde{\xi}_n > x + n\tilde{c}\}$. Then, from Veraverbeke's theorem,

$$\mathbf{P}(B) \leq \mathbf{P}(C) + \mathbf{P}(B \setminus C) \equiv \mathbf{P}(C) + o(\mathbf{P}(C)).$$

So

$$\begin{aligned} \mathbf{P}(A) \quad &= \quad \mathbf{P}(AB) \leq \mathbf{P}(AC) + \mathbf{P}(AB \setminus AC) \\ &= \quad \mathbf{P}(AC) + \mathbf{P}(A(B \setminus C)) \\ &\leq \quad \mathbf{P}(AC) + \mathbf{P}(B \setminus C) \\ &= \quad \mathbf{P}(AC) + o(\mathbf{P}(C)) \\ &\leq \quad \sum_n \mathbf{P}(AC_n) + o(\mathbf{P}(C)) \\ &= \quad \sum_n \mathbf{P}(AC_n) + o(\overline{F}^s(x)). \end{aligned}$$

We now return to the evaluation of Σ_1 and Σ_2 defined in (77). We have

$$\Sigma_1 \leq (1+o(1))\frac{\varepsilon d^{(2)} M}{\tilde{c}}\overline{F}^s(x).$$

On the intersection of the events $D_{n,\varepsilon,R}$, $S_{n-1} \leq R-(n-1)(c-\varepsilon)$, and $\{\xi_n \leq x-2R+(n-1)(c-\varepsilon)\}$, we have $S_n \leq x-R$. In addition, for all $j \geq 1$,

$$S_{n+j} = S_{n+j} + S_n - S_n \leq \tilde{S}_{n+j} - \tilde{S}_n + S_n \leq R - j(\tilde{c}-\varepsilon) + x - R \leq x.$$

So, when x is large enough, on this intersection, for all m $S_m < x$. Therefore

$$\begin{aligned}\mathbf{P}(D_{n,\varepsilon,R}, \overline{S} > x) &= \mathbf{P}(D_{n,\varepsilon,R}, \overline{S} > x, \xi_n \leq x-2R+(n-1)(c-\varepsilon)) \\ &\leq \mathbf{P}(\xi_n \leq x-2R+(n-1)(c-\varepsilon)).\end{aligned}$$

So

$$\begin{aligned}\Sigma_2 &\leq \sum_n \mathbf{P}(\xi_n > x-2R+(n-1)(c-\varepsilon)) \\ &= (1+o(1))\frac{1}{c-\varepsilon}\int_{x-2R}^{\infty}\mathbf{P}(\xi_1 > t)dt \\ &= (1+o(1))\frac{d^{(2)}M}{c-\varepsilon}\overline{F}^s(x-2R) \\ &= (1+o(1))\frac{d^{(2)}M}{c-\varepsilon}\overline{F}(x),\end{aligned}$$

because of (75). Since $\varepsilon > 0$ is arbitrary, the result follows. □

Case $b^{(1)} = b^{(2)}$ We assume $v_i^2 = \mathrm{var}\sigma^{(i)}$ to be finite for $i = 1, 2$ and we use the notation $v = \sqrt{v_1^2 + v_2^2}$.

Theorem 8 *Assume $a > b^{(1)} = b^{(2)} \equiv b$. Then, as $x \to \infty$,*

$$\mathbf{P}(W^{(2)} > x) \sim 2d^{(1)}\int_0^{\infty}\overline{F}(x+y(a-b))\overline{\Phi}\left(\frac{x}{v\sqrt{y}}\right)dy + \frac{d^{(2)}M}{a-b}\overline{F}^s(x), \qquad (78)$$

where $\overline{\Phi}$ is the tail of the standard normal distribution.

Proof Note that, for any $\varepsilon_n \to 0$,

$$\mathbf{P}(A_x^{(1)}) \sim \sum_{n=0}^{\infty}\mathbf{P}(\sigma_{-n-1}^{(1)} > x + n(a-b+\varepsilon_n)).$$

Put

$$l \equiv l_n = \frac{n(a-b+\varepsilon_n)}{a}$$

(w.l.o.g. we can assume l to be an integer). Put

$$\tau \equiv \tau_n = \min\{m : \quad \sum_1^m (a - \sigma_{-n+i-1}^{(1)}) > la\}$$

(if $\sigma_{-n-1} > x + n(a-b+\varepsilon_n)$, then $-n+\tau_n$ is the index of the first customer after $-n$ who possibly finds station 1 empty. In particular, inter-arrival times to the second station are $\sigma^{(1)}$'s for all customers between $-n$ and $-n+\tau_n$. From the SLLN, there exists a sequence $1 \geq \delta_n \to 0$ such that, for all n,

$$\mathbf{P}\left(\left|\frac{\tau_n}{la} - \frac{1}{a-b}\right| > \delta_n\right) \leq \delta_n.$$

Therefore,

$$\mathbf{P}(\tau_n \geq \frac{la}{a-b} - la\delta_n) \geq 1 - \delta_n.$$

Substitute n:

$$\frac{la}{a-b} - la\delta_n \quad = \quad n + \frac{n\varepsilon_n}{a-b} - la\delta_n \geq n + \frac{n\varepsilon_n}{a-b} - n(a-b+1)\delta_n.$$

Take

$$\varepsilon_n = 2\delta_n(a-b)(a-b+1).$$

The RHS of the latter inequality is not less than n.

So, given the event $\{\sigma_{-n-1} > x + n(a-b+\varepsilon_n)\}$, with probability at least $1-\delta_n$, all n customers with numbers $-n, -n-1, \ldots, -1$ will be served at station 1 without interruptions.

Consider now a GI/GI/1/∞ queue with interarrival times $\{\sigma_i^{(1)}\}$ and service times $\{\sigma_i^{(2)}\}$ with the same mean $b^{(1)} = b^{(2)} = b$. Denote by W_n the waiting time of customer n. Then

$$\frac{W_n}{v\sqrt{n}} \to \eta$$

weakly, where η has the tail distribution

$$\begin{aligned} \mathbf{P}(\eta > x) &= 2\overline{\Phi}(x), \\ \overline{\Phi}(x) &= \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\{-y^2/2\} dy. \end{aligned}$$

Take any $c > 0$. If $x \to \infty$, $N_x \leq n \leq cx^2$ (where N_x is as defined in the proof of Theorem 6), then

$$\mathbf{P}\left(\frac{W_n}{v\sqrt{n}} > \frac{x}{v\sqrt{n}}\right) \quad \leq \quad \mathbf{P}\left(\frac{W_n}{v\sqrt{n}} > \frac{1}{v\sqrt{c}}\right) = (1+o(1))2\overline{\Phi}\left(\frac{1}{v\sqrt{c}}\right).$$

For any $\Delta > 0$, choose $c << 1$ such that $\overline{\Phi}(\frac{1}{v\sqrt{c}}) \leq \Delta$. Then

$$\sum_{N_x}^{cx^2} \mathbf{P}(\sigma_{-n-1}^{(2)} > x + nb, W^{((2)} > x) \leq \Delta \overline{F}^s(x).$$

If $n > cx^2$, then

$$\mathbf{P}\left(\frac{W_n}{v\sqrt{n}} > \frac{x}{v\sqrt{n}}\right) = (1+o(1))2\overline{\Phi}\left(\frac{x}{v\sqrt{n}}\right),$$

since

$$\mathbf{P}\left(\frac{W_n}{v\sqrt{n}} > y\right) = (1+o(1))2\overline{\Phi}(y)$$

uniformly in y on a compact set.

Therefore (we omit details concerning the fact that, before the first large jump, with a probability close to 1, everything lies in compact set),

$$\begin{aligned}
&\sum_{cx^2}^{\infty} \mathbf{P}(\sigma_{-n-1}^{(1)} > x + n(a-b), W^{(2)} > x) \\
= \;& (1+o(1)) \sum_{cx^2}^{\infty} \mathbf{P}(\sigma_{-n-1}^{(1)} > x + n(a-b+\varepsilon_n), W^{(2)} > x) + o(\overline{F}^s(x)) \\
= \;& (1+o(1)) \sum_{cx^2}^{\infty} \mathbf{P}(\sigma_{-n-1}^{(1)} > x + n(a-b+\varepsilon_n)) \mathbf{P}\left(\frac{W_n}{v\sqrt{n}} > \frac{x}{v\sqrt{n}}\right) + o(\overline{F}^s(x)) \\
= \;& 2\sum_{cx^2}^{\infty} \mathbf{P}(\sigma_{-n-1}^{(1)} > x + n(a-b)) \overline{\Phi}\left(\frac{x}{v\sqrt{n}}\right) + o(\overline{F}^s(x)) \\
= \;& 2\int_{cx^2}^{\infty} d^{(1)} \overline{F}(x + y(a-b)) \overline{\Phi}\left(\frac{x}{v\sqrt{y}}\right) dy + o(\overline{F}^s(x)) \\
= \;& 2d^{(1)} \int_{0}^{\infty} \overline{F}(x + y(a-b)) \overline{\Phi}\left(\frac{x}{v\sqrt{y}}\right) dy + o(\overline{F}^s(x)).
\end{aligned}$$

□

Comments Obviously, the term obtained is $O(\overline{F}^{(s)}(cx^2))$. Therefore, it is $o(\overline{F}^{(s)}(x))$, say, for Pareto and for Weibull distributions. However, if $\overline{F}^{(s)}$ is "extremely heavy", say

$$\overline{F}^{(s)}(x) \sim (\log x)^{-K}$$

for a certain positive constant K (note: it is SE !), then a term of order $O(\overline{F}^{(s)}(cx^2))$ is not negligible.

Case $b^{(1)} < b^{(2)}$ For any $c > 1$ and $\varepsilon \in (0, c-1)$, put

$$h(c,\varepsilon) = \limsup_{x\to\infty} \frac{F^s(c(1+\varepsilon)x) - F^s(c(1-\varepsilon)x)}{\overline{F}^s(x)}.$$

Theorem 9 *Assume $b^{(1)} < b^{(2)}$. Assume, in addition, the function F to be such that, for any $c > 1$,*

$$\lim_{\varepsilon\to 0} h(c,\varepsilon) = 0. \tag{79}$$

Then

$$\mathbf{P}(W^{(2)} > x) \;\sim\; \frac{d^{(2)}M}{a-b^{(2)}} \overline{F}^s(x) + \frac{d^{(1)}M}{a-b^{(2)}} \overline{F}^s\left(x\frac{a-b^{(1)}}{b^{(2)}-b^{(1)}}\right). \tag{80}$$

Proof We have to find the asymptotics for

$$P(x) \equiv \mathbf{P}(\{W_0^{(2)} > x\} \bigcap A_x^{(1)}) \equiv P_1(x) + P_2(x)$$

where

$$P_1(x) \quad = \quad (1+o(1)) \sum_{n=N_x}^{\infty} \mathbf{P}(W_{-n}^{(1)} \le K, W_{-n-1}^{(2)} \le K, A_{x,n}^{(1)}, W_0^{(2)} > x)$$

and

$$\begin{aligned} P_2(x) &\le (1+o(1)) \sum \mathbf{P}(\{W_{-n}^{(1)} > K\} \bigcup \{W_{-n-1}^{(2)} > K\}) \mathbf{P}(A_{x,n}^{(1)}) \\ &\le (1+o(1)) \sum \mathbf{P}(A_{x,n}^{(1)}) (\mathbf{P}(W_0^{(1)} > K) + \mathbf{P}(W_0^{(2)} > K)). \end{aligned}$$

For any $\delta > 0$, one can choose $K \equiv K(\delta)$ such that

$$P_2(x) \le (1+o(1)) \delta \mathbf{P}(A_x^{(1)}).$$

Put $n(x) = \frac{x}{b^{(2)} - b^{(1)}}$ and, for a sufficiently small $\varepsilon \in (0,1)$, $n_1(x) = n(x)(1-\varepsilon)$, $n_2(x) = n(x)(1+\varepsilon)$. Then

$$P_2(x) = \sum_{N_x}^{n_1(x)} + \sum_{n_1(x)}^{n_2(x)} + \sum_{n_2(x)}^{\infty} \equiv P_3(x) + P_4(x) + P_5(x),$$

where

$$\begin{aligned} P_4(x) &\le (1+o(1)) \mathbf{P}(\bigcup_{n_1(x)}^{n_2(x)} A_{x,n}^{(1)}) \\ &= (1+o(1)) \frac{F^s(x + n_1(x)a) - F^s(x + n_2(x)a)}{a-b}. \end{aligned}$$

From (79), for any $\delta > 0$, one can choose $\varepsilon \in (0, (b^{(2)} - b^{(1)})^{-1})$ such that

$$P_4(x) \le (1+o(1)) \delta \overline{F}^s(x).$$

Further,

$$P_3(x) \quad \le \quad (1+o(1)) \sum_{N_x}^{n_1(x)} \mathbf{P}(A_{x,n}^{(1)}) \mathbf{P}(K + \max_{1 \le j \le n} \sum_{i=1}^{j} (\sigma_{-i}^{(2)} - \sigma_{-i-1}^{(1)}) > x).$$

Put

$$V_n = \max_{1 \le j \le n} \sum_{i=1}^{j} (\sigma_{-i}^{(2)} - \sigma_{-i-1}^{(1)}).$$

Then

$$\frac{V_n}{n} \to b^{(2)} - b^{(1)}$$

a.s. and

$$\mathbf{P}(V_{n_1(x)} > x - K) \to 0$$

for any fixed K. Therefore,

$$P_3(x) \quad \le \quad (1+o(1)) \mathbf{P}(V_{n_1(x)} > x - K) \mathbf{P}(A_x^{(1)}) = o(\overline{F}^s(x)),$$

as $x \to \infty$. Thus,

$$P(x) = (1 + o(1))P_5(x) + o(\overline{F}^s(x)).$$

Consider $P_5(x)$. For any $\Delta > 0$, put

$$\begin{aligned} P_5(x) &= (1 + o(1)) \sum_{n_2(x)}^{\infty} \mathbf{P}(W_0^{(2)} > x, W_{-n}^{(1)} \le K, W_{-n-1}^{(2)} \le K, \sigma_{-n}^{(1)} > x + n(a - b + \Delta)) + P_7(x) \\ &\equiv P_6(x) + P_7(x), \end{aligned}$$

where, for any $\delta > 0$, one can choose Δ such that

$$P_7(x) \le (1 + o(1))\delta \overline{F}^s(x).$$

Consider $P_6(x)$. Note that a single-server queue is monotone in interarrival times. For any $n \ge n_2(x)$, on the event $\{\sigma_{-n} > x + n(a - b^{(2)} + \Delta)\}$ consider the minimal possible value $\sigma_{-n}^{(1)} \equiv y = x + n(a - b + \Delta)$. Then approximately y/a customers arrive to station 1 to the moment when a service of customer $-n$ is completed. Denote by $y/a + l(y) - n$ a number of the first (after $-n$) customer who finds station 1 empty. Due to the SLLN,

$$\frac{y}{a} b^{(1)} + l(y) b^{(1)} = (1 + o(1)) l(y) a$$

as $n \to \infty$ and

$$\frac{y}{a} + l(y)(1 + o(1)) \frac{y}{a - b^{(1)}} < n$$

if Δ is sufficiently small. Then, with probability close to 1,

$$\begin{aligned} W_0^{(2)} &\ge \frac{y}{a - b^{(1)}} (b^{(2)} - b^{(1)}) - \left(n - \frac{y}{a - b^{(1)}} \right) + o(n) \\ &= -n(a - b^{(2)}) + y + o(n) \\ &= x + n\Delta - o(n) \ge x. \end{aligned}$$

Thus,

$$|P_5(x) - \sum_{n=n_2(x)}^{\infty} \mathbf{P}(A_{x,n}^{(1)})| \le (1 + o(1))\delta \overline{F}^s(x).$$

Since $\delta > 0$ is arbitrary, the result follows. □

5.3.3 Multiserver Queues

The aim of this section is to derive upper and lower bounds for the asymptotic tail behavior of the steady state maximal dater of multiserver queues. Since (AA) does not hold, we cannot use the approach of §4. We show how the ideas of §5.2 can be used to derive upper and lower bounds which are specific to this queue.

Thanks to the results of §6.1, we can replace the GI/GI/m/∞ queue by a D/GI/m/∞ queue with constant inter-arrival times a. Let $\mathbf{E}\sigma = b$ and $\rho \equiv \frac{b}{ma} \in (0, 1)$. Assume further that $\mathbf{P}(\sigma_1 > x) = \overline{F}(x)$, where both distributions F and F^s are sub-exponential.

Theorem 10 *Under the foregoing assumptions, when x tends to ∞,*

$$\mathbf{P}(Z > x) \quad = \quad (1 + o(1)) \left(m\rho \overline{F}^s(x) + \left(\frac{\rho}{1-\rho} - m\rho \right)^+ \overline{F}^s \left(\frac{bx}{b - (m-1)a} \right) \right). \qquad (81)$$

Note that the inequalities $\frac{\rho}{1-\rho} > m\rho$ and $b > (m-1)a$ are equivalent, and the second term in the RHS of Equation (81) disappears when $b \leq (m-1)a$.

The proof consists of three steps:

- first, we get a lower bound by using the SLLN;
- then we get an upper bound by using results from Section 4.3;
- finally, Theorem 4 gives us the tool to derive the exact asymptotics.

Lower Bound Clearly,

$$\mathbf{P}(Z > x) \geq \mathbf{P} \left(\bigcup_{n=0}^{\infty} \{\sigma_{-n} > x + na\} \right).$$

Due to reasons explained in Section 5.1, the occurrence of two or more big jumps is unlikely and

$$\begin{aligned} \mathbf{P}(Z > x) \quad &\geq \quad (1 + o(1)) \sum_{n=0}^{\infty} \mathbf{P}(\sigma_{-n} > x + na) \\ &= \quad (1 + o(1)) \frac{1}{a} \int_x^{\infty} \overline{F}(y) dy \\ &= \quad (1 + o(1)) \frac{b}{a} \overline{F}^s(x). \end{aligned}$$

Upper Bound Take a sufficiently large L and consider the L-upper-bound D/GI/1/∞ queue with inter-arrival times La and service times $\{\hat{s}_n\}$ with mean $\hat{b} = \mathbf{E}\hat{s}_1$. Since

$$\max_{1 \leq i \leq L} \sigma_i \leq \hat{s}_1 \leq \sum_1^L \sigma_i,$$

we get

$$\mathbf{P}(\hat{s}_1 > x) \sim L\mathbf{P}(\sigma_1 > x) = L\overline{F}(x)$$

as $x \to \infty$. Note that, for the multi-server queue, $\lambda = 1/a$ and $\gamma(0) = b/m$. Therefore, we get a natural analogue of Lemma 9:

$$\limsup_{x \to \infty} \frac{\mathbf{P}(Z > x)}{\overline{F}^s(x)} \quad \leq \quad \limsup_{x \to \infty} \frac{\mathbf{P}(\hat{R} > x)}{\overline{F}^s(x)} = \frac{Lb}{La - \hat{b}} \overline{F}^s(x). \qquad (82)$$

Thus, we are in a position to make use of Theorem 4. The rest of the proof of the theorem is omitted.

6 Appendix

6.1 Deterministic Interarrival Times

We extend to the monotone separable framework the approach used in [1] for single server queues to show that there may be no loss of generality in assuming that a network has deterministic interarrival times when one wants to evaluate the tails asymptotics of its stationary maximal dater.

The framework is that of Section 2. Fix $\{\xi_n\}$, $\{f_l\}$ and consider a family of networks with different 'input sequences' $\{T_n\}$ such that $\mathbf{E}\tau_1 > \gamma(0)$. W.l.o.g. assume $T_0 = 0$.

In particular, a network with constant inter-arrival times (say a) belongs to this family. We will supply its characteristics by the upper index (a).

For any $\{T_n\}$ and for any $d < \mathbf{E}\tau_1$, set

$$\psi(\{\tau_n\}, d) = \sup_{n\geq 0}(nd + T_{-n}) \equiv \sup_{n\geq 0}\left(\sum_{i=-n}^{-1}(d - \tau_i)\right).$$

Theorem 11 *Assume that there exist a continuous and strictly positive function $h : (\gamma(0), \infty) \to (0, \infty)$ and a long-tailed distribution G such that, for any $a > \gamma(0)$,*

$$\mathbf{P}(Z^{(a)} > x) \sim h(a)\overline{G}(x) \quad as \quad x \to \infty. \tag{83}$$

Then, for any network with random interarrival times $\{\tau_n\}$, such that $\mathbf{E}\tau_1 = a > \gamma(0)$, the following is valid: if $\{\tau_n\}$ and $\{\xi_n\}$ are independent and if, for any $d < a$,

$$\mathbf{P}(\psi(\{\tau_n\}, d) > x) = o(\overline{G}(x)) \quad as \quad x \to \infty, \tag{84}$$

then

$$\mathbf{P}(Z > x) \sim h(a)\overline{G}(x) \quad as \quad x \to \infty. \tag{85}$$

Remark 7 *In particular, condition (84) is satisfied if the τ_n's are i.i.d. Indeed, then $\psi(\{\tau_n\}, d)$ has either a bounded (if $\mathbf{P}(d > \tau_1) = 0$) or exponential tail, which is lighter than any long tail.*

Proof Take any $\varepsilon \in (0, a - \gamma(0))$. Due to the monotonicity,

$$Z_{[-n,0]} \leq Z^{(a-\varepsilon)}_{[-n,0]} + \max_{0\leq j\leq n}\left(\sum_{i=-j}^{-1}(a - \varepsilon - \tau_i)\right)^+.$$

Therefore,

$$Z \leq Z^{(a-\varepsilon)} + \psi(\{\tau_n\}, a - \varepsilon) \equiv Z^{(a-\varepsilon)} + \psi$$

where $Z^{(a-\varepsilon)}$ and ψ are independent. Therefore,

$$\mathbf{P}(Z > x) \leq \mathbf{P}(Z^{(a-\varepsilon)} + \psi > x) \sim \mathbf{P}(Z^{(a-\varepsilon)} > x) \sim h(a - \varepsilon)\overline{G}(x).$$

Thus,

$$\limsup_{x\to\infty} \frac{\mathbf{P}(Z > x)}{\overline{G}(x)} \leq h(a - \varepsilon)$$

for any $\varepsilon \in (0, a - \gamma(0))$. letting ε go to 0, we get the upper bound $h(a)$.

For the lower bound, we use the monotonicity, the SLLN for the τ's, the LT and the independence assumptions. For any $\varepsilon > 0$, one can choose a sufficiently large C such that

$$\mathbf{P}(T_{-n} \geq -n(a+\varepsilon) - C \quad \forall n \geq 0) \geq 1 - \varepsilon.$$

Denote the latter event by D_ε. Then

$$\begin{aligned} & \mathbf{P}(Z > x) \geq \mathbf{P}(Z > x, D_\varepsilon) \geq \mathbf{P}(Z^{(a+\varepsilon)} - C > x, D_\varepsilon) \\ \geq \; & \mathbf{P}(Z^{(a+\varepsilon)} - C > x)(1-\varepsilon) \sim h(a+\varepsilon)(1-\varepsilon)\overline{G}(x+C) \sim h(a+\varepsilon)(1-\varepsilon)\overline{G}(x). \end{aligned}$$

Thus, for any $\varepsilon \in (0,1)$,

$$\liminf_{x\to\infty} \frac{\mathbf{P}(Z > x)}{\overline{G}(x)} \geq h(a+\varepsilon)(1-\varepsilon).$$

Letting ε go to 0, we get the lower bound with coincides with the upper one. □

6.2 Proof of Lemma 11

W.l.o.g., we prove the result for $j_1 = 1$, $j_2 = r$. Note that

$$\frac{\mathbf{P}(\sum_1^r Y_1^{(j)} > x \mid \nu)}{\overline{F}(x)} \to \sum_1^r d_\nu^{(j)} \leftarrow \frac{\mathbf{P}(\max_j Y_1^{(j)} > x \mid \nu)}{\overline{F}(x)}$$

$\mathbf{P}_\nu$-a.s. and, for all x,

$$\frac{\mathbf{P}(\sum_1^r Y_1^{(j)} > x \mid \nu)}{\overline{F}(x)} \leq \prod_1^r \tilde{d}_\nu^{(j)} \cdot \sup_x \frac{\overline{F}^{*r}(x)}{\overline{F}(x)}$$

where the latter supremum is finite. Then the Dominated Convergence Theorem implies that

$$\frac{\mathbf{P}(\sum_1^r Y^{(j)} > x)}{\overline{F}(x)} \equiv \mathbf{E}\left(\frac{\mathbf{P}(\sum_1^r Y_1^{(j)} > x \mid \nu)}{\overline{F}(x)}\right) \to d \equiv \sum_j \mathbf{E} d_\nu^{(j)} \equiv \sum_j d^{(j)}$$

and

$$\frac{\mathbf{P}(\max_j Y_1^{(j)} > x)}{\overline{F}(x)} \to d.$$

References

[1] S. Asmussen, H.P. Schmidli and V. Schmidt (1999), Tail Probabilities for Nonstandard Risk and Queueing Processes, *Stochastic Processes and Appl.*, 79, 265-286.

[2] F. Baccelli and S. Foss (1994), Ergodicity of Jackson-type Queueing Networks, *Queueing Systems*, 17, 5-72.

[3] F. Baccelli and S. Foss (1995), On the Saturation Rule for the Stability of Queues. *J. Appl. Prob.*, 32, 494-507.

[4] F. Baccelli and S. Foss, Tails in Generalized Jackson Networks with Subexponential Distributions. In preparation.

[5] F. Baccelli, A. Makowski and A. Shwartz (1989), The Fork Join Queue and Related Systems with Synchronisation Constraints : Stochastic Dominance, Approximations and Computable Bounds. *Adv. Appl. Prob.*, Vol. 21, No. 3, pp. 629-660.

[6] F. Baccelli, S. Schlegel, and V. Schmidt (1999), Asymptotics of Stochastic Networks with Sub-exponential Service Times. *Queueing Systems*, 33, 205-232.

[7] O.J. Boxma, Q. Deng, and A.P. Zwart (1999), Waiting-time Asymptotics for the M/G/2 Queue with Heterogeneous Servers. Memorandum COSOR 99-20, Eindhoven University of Technology.

[8] O.J. Boxma, Q. Deng (2000). Asymptotics behavious of the tandem queueing system with identical service times at both queues, *Math.Methods in Oper.Res.*, 52, 307-323.

[9] P. Embrechts, C. Kluppelberg, Th. Mikosch (1997) Modeling Extremal Events, Springer Verlag.

[10] S. Foss (1980), Approximation of Multichannel Queueing Systems. *Siberian Math. J.*, 21, No.6, 132-140.

[11] S. Foss and D. Korshunov (2000), Sampling at a Random Time with a Heavy-tailed Distribution. *Markov Processes and Related Fields*, No. 6, 534-568.

[12] T. Huang and K. Sigman (1999), Steady State Asymptotics for Tandem, Split-Match and other Feedforward Queues with Heavy Tailed Service. *Queueing Systems*, 33, 233-259.

[13] Queueing Systems (1999), Special Issue 33.

[14] A. Scheller-Wolf and K. Sigman (1997), Delay Moments for FIFO GI/GI/c Queues. *Queueing Systems*, 25, 77-95.

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38330 Montbonnot-St-Martin (France)
Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399